MOTION ESTIMATION FOR SUPER-RESOLUTION BASED ON RECOGNITION OF ERROR ARTIFACTS

Ana Stojkovic, Zoran Ivanovski

University Ss Cyril and Methodius, Skopje, Macedonia

ABSTRACT

The work presents an effective approach for subpixel motion estimation for Super-resolution (SR). The objective is to improve the quality of the estimated SR image by increasing the accuracy of the motion vectors used in the SR procedure. The correction of the motion vectors is based on appearance of error artifacts in the SR image, introduced due to registration errors. First, SR is performed using full pixel accuracy motion vectors obtained using full search block matching algorithm (FS-BMA). Then, machine learning based method is applied on the resulting images in order to detect and classify artifacts introduced due to missing subpixel components of the motion vectors. The outcome of the classification is a subpixel component of the motion vector. In the final step, SR process is repeated using the corrected (subpixel accuracy) motion vectors.

Index Terms *— super-resolution, image registration, machine learning, artifacts detection*

1. INTRODUCTION

The importance of the accuracy of the registration for successful multiframe SR is well known. Therefore, considerable attention has been put to registration adapted to multiframe SR procedure. The approach that is usually utilized in combination with iterative SR algorithms is successive improvement of the accuracy, with every iteration of the SR procedure [1-6]. This approach, although computationally expensive, does not guarantee accurate registration.

Most of the multiframe SR algorithms perform registration prior to SR using well established methods [7-8]. Often used are block-matching algorithms (BMA) [7], among which full-search (FS) is the most accurate and the most complex. Nevertheless, errors in estimated motion vectors (MVs) are likely to occur due to multiple local minima, presence of noise, etc.

Some of the approaches use methods from statistical-learning theory [9].

In [10] an approach for fast sub-pixel motion estimation used in video coding was proposed. It avoids interpolation and utilizes Taylor approximation and BMA with non overlapping blocks. The approach has reportedly good results; therefore we have included it in our comparison set, appropriately modified to be used in SR. In terms of SR, Vandewalle et al. in [11] proposed an approach for registration of translational shifts and the angle of rotation between LR images. For estimating only translational global motion vector (GMV) they use phase correlation between spectra of different LR images on same globally moving scene, and find the best solution using least squares method. Their algorithm performs well in cases of higher presence of edges. With images with lower spatial activity the algorithm performs poor. Another iterative planar motion estimation algorithm based on Taylor expansions, that gives similar results as [11], and performs in spatial domain, was proposed by Keren et al, [12]. Our tests show that this algorithm outperforms [11] in cases of lower amount of edges of some strong directionality.

In order to cope with the registration errors, SR procedures usually incorporate some kind of mechanism for avoiding introduction of artifact in the SR image. To deal with outliers, some of the algorithms use norm L_1 minimization of the cost function, [13-14]. This approach is efficient in suppressing the influence of the outliers, but the resulting SR image is lacking in sharpness compared to the image obtained using norm L_2-based minimization [15]. The approach of suppression of artifacts due to registration errors has gained considerable attention. The suppression is usually performed using regularization with appropriate term that incorporates prior knowledge about the appearance of the artifacts [16-21]. The work of Kattsaggelos et al. [18-21] has considered significant attention in this area. Although some of the algorithms are quite efficient, in many cases the sharpness of the resulting image is also affected.

Effects of errors in motion vectors (MV) appear as very pronounced and annoying artifacts around edges of the objects in SR image. In [22] it was shown that artifacts due to miss-registration have a nature of ringing effect. They are mostly expressed around edges that are perpendicular to the error direction, and their intensity is proportional to the magnitude of the error vector (EV).

The objective of this work is utilization of this information in order to correct motion vectors. The proposed algorithm performs correction of motion vector based on recognition of error artifacts in SR image. The paper presents proof of the concept and focuses its attention to global translational motion.

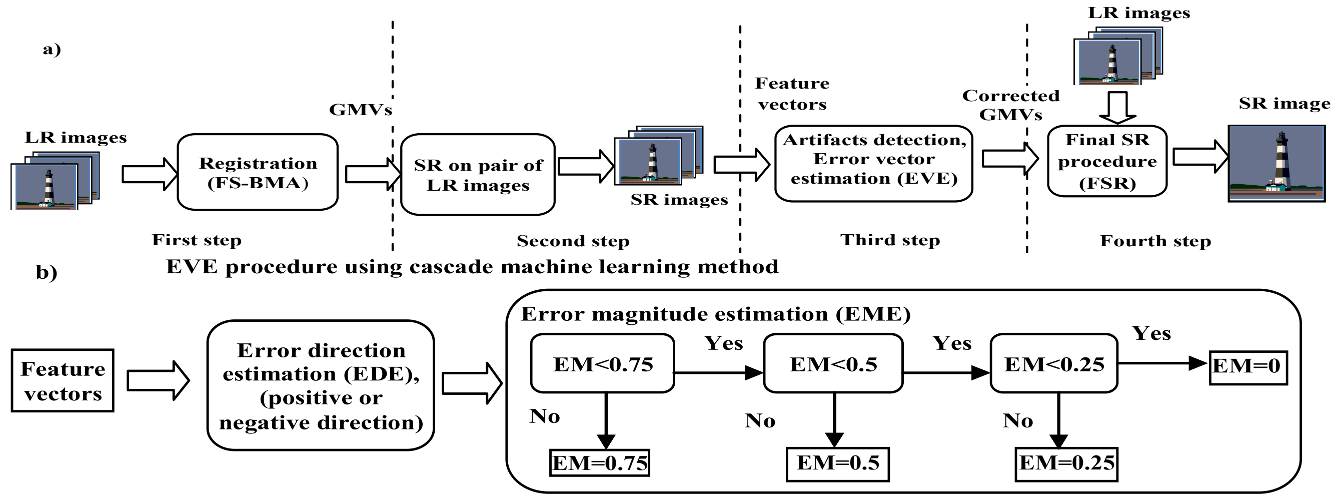

Fig 1 Block diagrams: proposed approach (a) and EVE procedure (b)

2. PROPOSED ALGORITHM

The algorithm represents a combination of motion estimation and SR. It consists of the following steps: initial motion estimation, SR with initially estimated motion vectors, detection and classification of error artifacts for correction of motion vectors, and final application of the SR procedure, now with accurate motion vectors. Block diagram of the proposed approach is shown in Fig.1- a) and b).

In the *first step*, registration is performed on pairs of images consisting of central image and one neighboring frame, using full pixel accuracy BMA-FS. Since the accuracy of the motion estimation is higher in the regions of the image with higher spatial activity, global vector for each pair of images is computed by averaging MVs in the regions around edges in the image. Then, in the *second step*, the calculated GMVs are used to perform super-resolution on every pair of images separately. Since the accuracy of the calculated GMVs is full pixel and the real motion in the images is rarely so, it is to be expected that pronounced artifacts will appear in every SR image.

In the *third step*, in order to find the magnitude and the direction of the error (subpixel component) of GMVs, by using information from artifacts appearance, cascade machine learning based method is applied for error vector estimation (EVE). The error vectors are estimated with quarter pixel accuracy and then added to the firstly estimated GMVs, in order to correct them.

In the *fourth step* the final super-resolution (FSR) is applied with use of all LR images, and the corrected global motion vectors.

In the following we explain in details the proposed procedure.

2.1. SR application on pairs of images

This step is applied in order to generate SR image containing artifacts introduced due to motion estimation inaccuracy. The artifacts should be pronounced, which would enable their efficient detection and recognition.

For every pair of LR images SR image is estimated using norm L_2-based minimization function. This type of minimization has good sharpening properties, but also high sensitivity to motion estimation errors, and produces SR image containing pronounced and easily distinguishable error artifacts. In order to further boost the artifacts, no regularization is used. The number of iterations highly affects the visibility of the artifacts. During the testing it was empirically established that 3 to 5 iterations are resulting in detectable presence of artifacts, not pronouncing the ringing artifacts due to other causes too much.

2.2. Artifacts detection and error vector estimation

Artifacts detection procedure is applied only on the regions of the image in the vicinity of the edges, since in those regions the appearance of the artifacts is to be expected. The extraction of those regions is performed using simple gradient threshold technique applied on the bilinearly interpolated image obtained from the central frame.

The detection and classification procedure can be divided into two steps. In the first step error direction estimation (EDE) is performed along each direction (horizontal and vertical) separately. The outcome of this estimation is positive or negative direction of the error vector component. EDE is sensitive on the transition of the luminance level. For that reason, prior to EDE, edges are classified into two categories, i.e. edges that have transition from lower to higher luminance level, and edges that have transition from higher to lower luminance level. Following that, two separate classifiers for this step are trained.

In the second step, relying on the knowledge about the direction of the error (positive or negative), magnitude of each EV component is estimated (error magnitude estimation - EME). Block diagram of the proposed EVE procedure is shown in Fig.1-b). The decision is made through cascade of binary classifiers, each of which decides on possible

0.98	0.96	0.97	0.97	0.86	0.81	0.91	0.83	0.95	0.96	0.92	0.96	0.89	0.89	0.89	0.89	0.89	0.85	0.88	0.88	0.92	0.81	0.96	0.92	0.81	0.77
FA	SA	FA	SA	FA	SA	FA	SA	FA	SA	FA	SA	FA	SA	FA	SA	FA	SA	FA	SA	FA	SA	FA	SA	FA	SA
Positive direction (PD)		Negative direction (ND)		PD		ND		PD		ND		PD		ND		PD		ND		PD		ND		0	
				0.75				<0.75				0.5				<0.5				0.25					
EDE acc. (positive or negative)				EME acc.(amp=0.75 or amp.<0.75)								EME acc. (amp=0.5 or amp.<0.5)								EME acc.(amp=0.25 or amp.=0)					

Table 1 EVE accuracy

magnitude of the error component. This is performed maximum three times, yielding the error vector with one of the following values: 0.75, 0.5, 0.25 and 0 pixels (quarter pixel precision of MVs).

Logistic regression is used as machine learning based method in each step and level. Only pixels in edge regions are considered in the procedure, and, accordingly, feature descriptors are extracted only for those pixels.

2.2.1 Feature vector extraction

Feature descriptor for each step and level is one dimensional luminance vector consisting of two parts. First part of the vector bears information from bilinear interpolation of the CLR, and the second incorporates information from the produced SR image. The length of the descriptor for EDE is 6 elements, 3 from each image centered on every pixel in the edge regions of the image. For the EME the descriptor is 10 pixels long, 5 pixels from each image.

Separate descriptors are calculated in both directions, horizontal and vertical.

2.2.2 Classification method

Inspired by results in [22], logistic regression as machine learning based method was used as a binary classifier, both for EDE and EME procedures. We have used logistic regression function similar to one used in [22]. Two versions of the algorithm were tested. In the first version of the algorithm (FA) logistic regression is computed for each pixel of the area around edges. The outcome of the regression is compared to threshold, determined in training process by applying grid search, and the decisions are stored. The final decision is made by majority of the decisions.

In the second version of the algorithm (SA), the final feature vector (FFV) is obtained by averaging feature vectors extracted from the regions around the edges. The final feature vector is used in a single logistic regression at any level of both EDE and EME procedures.

2.2.3. Training process

Training set for estimation of logistic regression of each level of EDE and EME is chosen to be 66% from all grayscale images included in both training and testing processes. Total number of used images is 153, out of which 102 are used in training process, and the rest 51 are used in the testing process. The images were created by cropping a part of 128x128 pixels from images obtained using 12 Mpixel digital camera. For the purposes in our experiments, LR images were obtained by downsampling each image with factor of 2. Neighboring LR images are obtained from every image in the training and testing process, by translation with all combinations of values with quarter pixel precision for the horizontal and vertical components of the global motion vector.

2.2.4. Performance testing of the detection procedure

In the testing process all combinations of error vector components that, when added to the real GMV, result in erroneous vector with full pixel precision (integer erroneous vector), are considered with same probability of appearance.

In Table 1 the accuracy of the binary classifiers is presented, in both variations of the algorithm, FA and SA, respectively.

The EDE part of the table represents the accuracy of the decisions concerning the direction (positive or negative) of the particular component (horizontal or vertical) of the error vector. In EME part of the table the accuracy of the decisions concerning the magnitude of the error is presented. Considered error magnitudes are 0.75, 0.5, 0.25 and 0 pixels. Certain table parts present the accuracy of the decision the error magnitude to be 0.75 pixels or less, 0.5 pixels or less and 0.25 pixels or 0. The presented accuracy values show that the proposed approach is high accurate in error vector estimation.

3. RESULTS

To test the performance of the complete algorithm 13 video sequences were taken. In order to compute PSNR, frames were downscaled with factor of 2, and then the proposed approach with FA and SA, with utilization of 5 neighboring frames was applied. For comparison, superresolution was performed using norm L_2-based minimization and GMVs estimated using different motion estimation (ME) algorithms: FS-BMA, algorithms proposed in [10], [11], and [12]. In all cases GMVs were quantized to quarter-pixel accuracy. The first 11 sequences were obtained from images recorded with 12Mpixel digital camera, by cropping a small part of 128x128. Last two sequences were extracted from HD (720x1280) documentary movie. Global translational motion was assumed in all sequences.

Sequences were chosen to be of different quality. The first, the tenth and the eleventh sequence are of poor quality, i.e. they contain a large amount of noise and compression artifacts. The rest of the sequences are of relatively good quality. Results from comparison in terms of Peak Signal to Noise Ratio (PSNR) and Mean structural similarity index (MSSIM) are given in graphics presented in Fig.2. From Fig.2 it can be seen that in terms of quality, our algorithm outperforms [10], and achieves comparable results to FS-BMA, [11] and [12], in most of the cases.

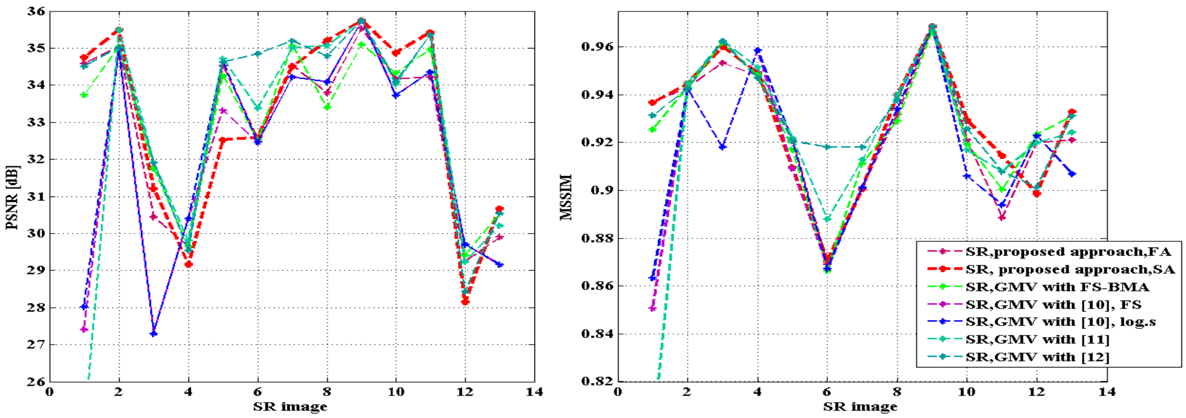

Fig 2 PSNR and MSSIM values for SR images achieved with different ME algorithms

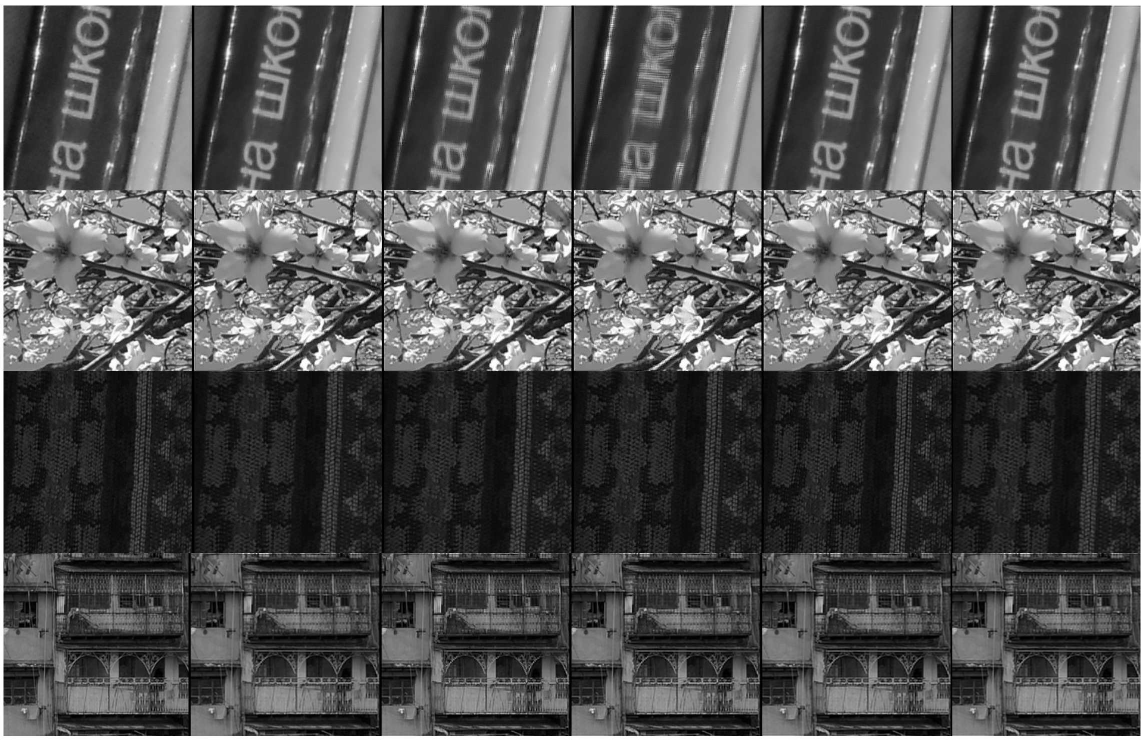

Fig 3 SR 1-4 a) original image; b) BMA-FS; c) GMVs from [10] FS; d) GMVs from [11]; e) GMVs from [12]; f) proposed approach SA

In order to enable visual comparison of the results, SR images from sequences numbered with 1, 4, 6, 13, obtained using different ME algorithms and same SR algorithm, are shown in Fig.3.1-4. The reason for not including the images obtained with FA version of the proposed algorithm is in their visual equivalence with the results obtained with SA.

For sequence 1 shown in Fig.3.1, in terms of ME, we achieve comparable results to those obtained with FS-BMA and with algorithm described in [12], and better results than those obtained with [10] and [11]. This result shows that the proposed algorithm outperforms those obtained with [10] and [11] for images lacking in edges. For the sequences presented in Fig.3.2 (natural scene) and Fig.3.4 (HD sequence 13), proposed approach performs similar to algorithms described in [10, 11 and 12]. For the sequence 6 shown in Fig.3.3, [11] and [12] outperform the proposed approach (in both FA and SA), because of the nature of edges. Texture in this image has a nature of well known ringing effect, which existence detection was objective in artifacts detection procedure. However, existing artifacts combined with the texture, do not cause unpleasant experience. The complete results can be seen here: http://dipteam.feit.ukim.edu.mk/?q=node/358.

The performance of the algorithms in PSNR (Fig.2) is very different, rendering this measure as uninformative. Nevertheless, it can be seen that the proposed approach achieves comparable performance in quality to the other tested algorithms. The performance of the algorithm in MSSIM (Fig.2) is good, which shows that the remaining artifacts in final SR image are visually neglectable.

In terms of computational complexity, general conclusion is that the proposed approach, especially in his SA version, has significantly lower computational complexity compared to quarter-pixel accuracy BMA, and has comparable complexity to algorithms proposed in [11] and [12].

4. CONCLUSION AND FUTURE WORK

The paper presents a new approach for super-resolution motion estimation based on error artifacts recognition using machine learning based methods. Compared to well established algorithms that estimate global translational motion between LR images, the proposed approach demonstrates comparable performances, visually as well as computationally.

Currently, the adaptation of the algorithm to motion estimation on a local level is considered, as well as its extension with rotational and perspective motion**.**

5. REFERENCES

[1] R.C.Hardie, K.J.Barnard, J.G.Bognar, E.E.Armstrong, and E.A.Watson, "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," vol.37, no.1,pp.247-260,Jan.1998.

[2] R. C. Hardie, K. J. Barnard and E. E. Armstrong, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images", *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1621-1633, December 1997.

[3] R. C. Hardie, T. R. Tuinstra, J. Bognar, K. J. Barnard, and E. E. Armstrong, "High Resolution Image Reconstruction From Digital Video with Global and Non-Global Scene Motion," Proceeding of the *IEEE Int. Conf. on Image Processing* (ICIP), Santa Barbara, CA, Oct. 1997.

[4] T. R. Tuinstra and R.C.Hardie, "High Resolution Image Reconstruction of Digital Video with Non-global, In-Scene Rigid Body Motion" Proceedings of *SPIE's International Symposium on Aerospace/Defense Sensing, Simulation, and Controls* (Aerosense), Orlando FL, April 1998

[5] S. Cain, R. C. Hardie and E. A. Armstrong, "Restoration of Aliased Video Sequences via a Maximum-Likelihood Approach," Proc. of the 1996 *Meeting of the IRIS Specialty Group on Passive Sensors*, Monterey CA, March 1996

[6] E. Kaltenbacher and R. C. Hardie, "High Resolution Infrared Image Reconstruction Using Multiple Low Resolution Aliased Frames," Proceedings of *IEEE National Aerospace Electronics Conference* (NAECON), Dayton, Ohio, May 1996

[7] Aroh Barjatya,"Block Matching Algorithms for Motion Estimation", DIP 6620 Spring 2004 Final Project Paper

[8] B.Lucas and T.Kanade, " An iterative image registration techique with an application to stereo vision". In Proc. DARPA Image Understanding Workshop, 1981, pp.121-130.

[9] H.Wechsler, Z.Duric, F.Li, V.Cherkassky, "Motion Estimation Using Statistical Learning Theory", IEEE Trans. on Pattern Rec. and Machine Intelligence, vol.26, no.4, April 2004.

[10] S.H.Chan, D.T.Võ, T.Q.Nguyen, "Subpixel motion estimation without interpolation", IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), March 2010, pp. 722 – 725

[11] P.Vandewalle, S.Süsstrunk, M.Vetterli, "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution ", EURASIP Journal on Applied Signal Processing, May 2005, vol.2006, pp.1-14.

[12] D.Keren, S.Peleg, and R.Brada, "Image sequence enhancement using sub-pixel displacements", Proc. of IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVRP'88), pp.742-746, Ann Arbor, Mich, USA, June 1988.

[13] A.Zomet, A.Rav_Acha, and S.Peleg, "Robust super resolution", in Proc. Int. Conf. Computer Vision and Pattern Recognition, vol.1, Dec.2001, pp.645-650.

[14] S.Farsiu, M.D.Robinson, M.Elad and P.Milanfar, "Fast and Robust Multiframe Super Resolution", *IEEE Trans. Image Processing,* vol.13, no.10, October 2004, pp.1327-1344.

[15] Z. Ivanovski, Lj. Panovski and L. J. Karam, "Robust super-resolution based on pixel-level selectivity", *Proc. SPIE*, vol. 6077, pp. 80-87, Visual Communications and Image Processing – VCIP, January 2006

[16] M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646–1658, Dec. 1997.

[17] N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Processing*, vol. 10, pp. 573–583, Apr. 2001.

[18] D. Babacan, R. Molina, and A. Katsaggelos, "Variational Bayesian Blind Deconvolution Using a Total Variation Prior", *IEEE Trans. Image Processing*, vol. 18, no.1, pp.12-26, 2009.

[19] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Parameter Estimation in TV Image Restoration Using Variational Distribution Approximation", *IEEE Trans. Image Processing*, vol. 17, no. 3, 326-339, March 2008.

[20] S. Villena, M. Vega, R. Molina, and A. K. Katsaggelos,"Bayesian super-resolution image reconstruction using an l1 prior", Proc. of 6th Int. Symp. on Image and Signal Processing and Analysis (ISPA 2009), Image Processing and Analysis Track, pages 152-157,2009.

[21] S. Villena, M. Vega, D. Babacan, R. Molina, and A. Katsaggelos, "Using the Kullback-Leibler divergence to combine image priors in super-resolution image reconstruction", Proc. of IEEE Int. Conf. on Image Processing, pages 809-812, Hong-Kong Sept. 2010.

[22] A.Milchevski, Z.Ivanovski, B.Mustafa, "Machine learning based super-resolution algorithm robust to registration errors", IEEE Digital Signal Proc. and Signal Proc. Education Workshop (DSP/SPE), Sedona, AZ, Jan. 2011, pp.326-331