

Combined Iterative Back Projected-Maximum a Posteriori Technique for Reconstructing Low Resolution Surveillance Videos

Temitayo Fagbola¹, Olatunde Olabiyisi², Elijah Omidiora³

LAUTECH

¹temitayo.fagbola@fuoye.edu.ng, ²soolabiyisi@lautech.edu.ng, ³eoomidiora@lautech.edu.ng

Abstract

Terrorism, ongoing destruction of public facilities, attacks on human and the recurring global security insurgences have posed new opportunities for the sporadic influx and deployment of Video Surveillance Systems. Consequentially, videos captured by these systems are typically of low resolution. This challenge greatly accounts for the failure of most existing video-based face recognition systems. However, most existing promising solutions to this challenge are computationally very expensive and inefficient for restoring continuous variation region and suppressing blocky artifacts within reasonable time bound. In this paper, a hybrid Iterative Back Projection – Maximum a Posteriori Technique (IBP-MAP) for Pixel domain super resolution reconstruction of low quality surveillance video feeds was developed. L1, a sparse prior and Simultaneous Auto Regression, a non-sparse prior of Bayesian MAP were combined with Iterative back projection, a spatial domain method to realize appreciable denoising, edge and non-edge preserving properties for super-resolved high resolution video frames in a computationally-efficient manner. The performance of the developed hybrid IBP-MAP technique was evaluated using Peak Signal to Noise Ratio (PSNR) and Improvement in Signal to Noise Ratio (ISNR). Results obtained using the hybrid IBP-MAP technique show significant improvements over other existing techniques in terms of quantitative and visual qualities of the video sequences in a time-efficient manner.

Keywords: *Pixel-level-Back-projection, Low-Quality-Video, Maximum-a-Posteriori, Super-Resolution-Reconstruction.*

I. Introduction

The demand for emerging Video Surveillance Systems (VSS) is immensely increasing in various fields of security applications and beyond as large publicly accessible facilities and urban sites, including airports, train stations, power plants, banking halls, shopping malls, museums, parking garages and hotels, now increasingly deploy comprehensive VSS [1],[2]. The sporadic influx of VSS has now posed new opportunities for the development and proliferation of video-based Face Recognition Systems (FRS) [3], [4].

FRS tends to identify or verify one or more persons in a video frame from a video source by using a stored database of faces [5]-[7]. Surveillance applications based on face recognition are gaining increasing attention after the United States' 9/11 events and with the ongoing security threats [8], [9]. Areas aiming at increased individual safety relative to terrorist threats further extend the integration of face recognition into VSS development [10], [11].

However, videos captured by surveillance cameras are typically of low quality due to Low-Resolution (LR) [12]-[15]. This challenge has greatly degraded the performance and also led to the failure of most existing video-based FRS [16], [17] when implemented in practice [18]. Thus, the development of an efficient video-based FRS becomes considerably a more challenging problem domain which by implication requires an equally wide range of accurate and computationally-efficient video enhancement method for video-based FRS to be usable.

Video enhancement methods help enhance the biometric content of videos. To achieve this, Super Resolution (SR) can be used to consolidate the information in successive low resolution frames to generate the details of facial features of potential high resolution highly crucial for human recognition and further analysis [19], [20]. In this process, a low resolution frame is being up-sampled by recovering the missing high frequency details and degradations in the frame with the objective of constructing a high resolution frame. In this paper, a pixel level back-projected maximum a posteriori super resolution technique for low quality surveillance video feeds is developed to address the low resolution problem.

II. Related works

Weisheng *et al.* [21] incorporated adaptively the non-local iterative back projection (NLIBP) algorithm into the IBP process so that the reconstruction errors can be reduced during image enlargement. The IBP technique iteratively reconstructs a HR image from its blurred and downsampled LR counterpart. However, the conventional IBP methods often produce many jaggy and ringing artifacts because the reconstruction errors are back projected into the reconstructed image isotropically. One major drawback of this approach is that it suffers from non-edge guidance.

Feng *et al.* [22] developed a video super-resolution reconstruction based on sub-pixel registration and Iterative Back Projection. A SRR method based on a sliding window was proposed utilizing the movement information between frames in the low-resolution video to improve the spatial resolution of video. A registration algorithm based on a four-parameter transformation model through Taylor series expansion was proposed, using an iterative solving method as well as the Gaussian pyramid image model to estimate the movement parameters from coarseness to fine while the frames were reconstructed using an IBP algorithm.

Stefanos *et al.* [23] developed a class of SR algorithms based on the *Maximum A Posteriori* (MAP) framework. These algorithms utilize a new multichannel image prior model, along with the state-of-the-art single channel image prior and observation models. A hierarchical (two-level) Gaussian non-stationary version of the multichannel prior was also defined and utilized within the same framework. However, this approach is computationally expensive and paid no attention to the minimization of the reconstruction error incurred by the use of MAP. This makes the solution inefficient in restoring continuous variation region and suppressing block artifacts.

Feng *et al.* [24] utilized Wiener filtering image restoration algorithm to generate an error-parameter curve at different motion distance from which the motion distance of the motion point spread function (PSF) was estimated approximately. The super-resolution image was reconstructed through the IBP algorithm.

Vilenna *et al.* [25] developed a SRR technique based on the combination of TV prior and SAR prior of the Bayesian restoration technique. A unique approximation was obtained by finding the distribution on the HR image given the observations that minimize a linear convex combination of the Kullback-Leibler (KL) divergences associated with the posterior distribution for combining the priors. This solution suffers from the over-smoothing of non-edge and inner regions of images.

Rujul and Nita [26] proposed a super resolution technique using the IBP method combined with Canny Edge Detection and Gabor Filter with difference image. This IBP method minimized the error significantly by back projecting the error and the high frequency component, using Canny Edge Detection and Gabor Filter with difference image, of up-sampled LR images to gather more back projecting error iteratively. This approach is robust to noise with edge preservation. However, canny edge detection does not incorporate additional image prior model that guides and help realize better super resolved image. As a result, it is not reliable and efficient for highly degraded images.

Patel [27] developed a SRR technique by combining an IBP method with the edge preserving infinite symmetrical exponential filter (ISEF). Sequel to the fact that IBP can minimize the reconstruction error significantly in an iterative manner and give good result, it suffers from ringing and chessboard effects because error is back-projected without edge guidance. ISEF was used to provide edge-smoothing by adding high frequency information. The resultant technique improves the image's visual quality with very fine edge details. Since ISEF does not incorporate additional

image prior model that guides and help realize better super resolved image, it is not reliable and efficient for highly degraded images when implemented in practice.

III. Materials and method

The SRR problem and the developed PBM super resolution technique are discussed in this section.

A. The SRR Problem

Given L LR frames y_k , such that $k=1, \dots, L$, the SRR problem is to find an estimate of the HR frame as depicted in figure 1. A number of LR frames with facial information are consolidated together to generate a high resolution frame.

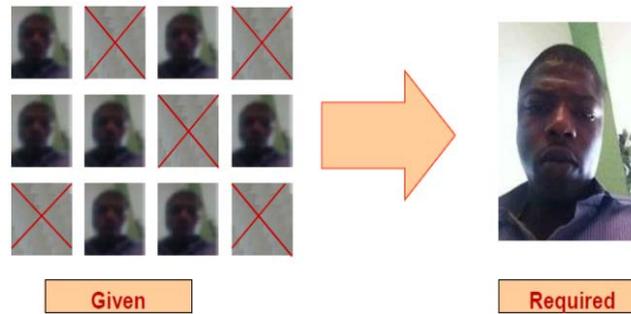


Fig.1. Resolution Reconstruction of Low Quality Frames.

The SRR imaging model is conceptually presented in figure 2 and is given by:

$$\{ \underline{Y}_k = A_k \mathbf{H}_k C(s_k) \underline{X} + \underline{n}_k \}_{k=1}^N$$

Where \underline{Y}_N is the measured LR images (noisy, blurry, down-sampled), \mathbf{H}_N is the blur can be extracted from the camera characteristics, A_N is the downsample dictated by the required resolution ratio, C_N is the warp that can be estimated using motion estimation s_N and \underline{n}_N is the additive noise.

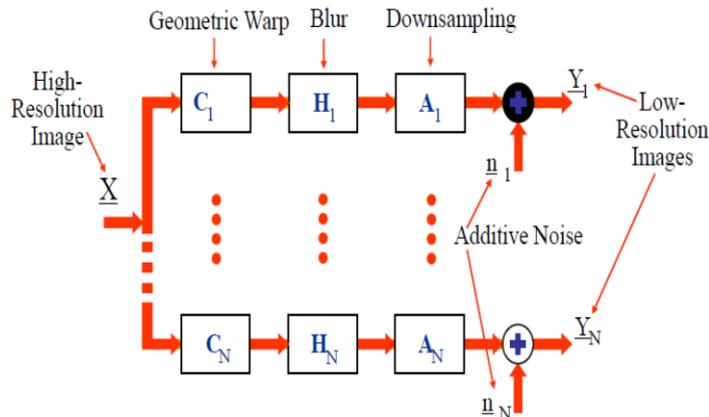


Fig. 2. The SRR Imaging Block Diagram.

B. The Developed PBM Super Resolution Technique

In this paper, a pixel level back-projected maximum a posteriori super resolution technique was developed using a combined l_1 - l_2 -SAR prior and iterative back projected technique. However, the super resolution reconstruction was performed through the minimization of the cost function defined by the l_2 norm regularization technique in equation (1)

$$Y(X) = \|\Gamma X\|_2 \tag{1}$$

where Γ is a high-pass operator that enhances both edge and noise equally. To model the unknown HR image x with the motion vectors $\{s_k\}$ using the proposed l_1 - l_2 -SAR prior depicted as:

$$P_{l_1-l_2-SARprior}(x|\alpha_{l_1-l_2-SAR}). \tag{2}$$

Prior distribution $p(s_k)$ was assigned to the unknown s_k , for $k = 1, \dots, L$ where L is the set of LR observations. The observation $y = \{y_k\}$ is a random process with the corresponding conditional distribution $p(y/x, \{s_k\}, \{\beta_k\})$. These distributions depend on additional parameters α_i and $\{\beta_k\}$ called hyperparameters which was modeled by assigning the hyperprior distributions. The conditional probability of the set of LR images y given x was expressed as an independent Gaussian model [28]:

$$p(y|x, \{s_k\}, \{\beta_k\}) = \prod_{k=1}^L p(y_k|x, s_k, \beta_k) \tag{3}$$

The hyperparameters $\alpha_{l,2}$ and $\{\beta_k\}$ are crucial for the performance of the SR algorithm, therefore, by using Gamma distributions to model the hyperparameters,

$$p(\omega) = \Gamma(\omega|\alpha_\omega^o, b_\omega^o) = \frac{b_\omega^o \alpha_\omega^o}{\Gamma \alpha_\omega^o} \omega^{\alpha_\omega^o - 1} \exp[-b_\omega^o \omega] \tag{4}$$

where $\omega > 0$ denotes a hyperparameter, $\alpha_\omega^o > 0$ and $b_\omega^o > 0$ are the shape and scale parameters respectively. The hyperpriors are chosen as Gamma distributions since they are conjugate priors for the Gaussian distribution. By implication,

$$p_l(\theta_{l,y}) = p_l(x|\alpha_l) p(\alpha_l) \prod_{k=1}^L [p(y_k|x, s_k, \beta_k) p(\beta_k) p(s_k)] \tag{5}$$

where $l \in \{1,2\}$ denotes the different joint probability distributions corresponding to the SAR, l_2 and l_1 prior models respectively.

Evaluated from Equation (5), the conditional probability for the set of L LR observations is

$$\theta_l = \{\Omega, \alpha_l\}, \Omega = \{x, \{s_k\}, \{\beta_k\}\}, p(\alpha_2) = p(\alpha_2^h) p(\alpha_2^v) \tag{6}$$

Let the set of all unknowns be denoted by:

$$\Phi = \{\Omega, \{\alpha_l\}\}, \tag{7}$$

Bayesian inference is based on the posterior distribution $p(\Phi|y)$ of Φ given the observed y . Since, the number of unknown pixels in HR grid X is usually very large, the HR image \hat{x} can be estimated such that when degraded it minimizes $p(AH_k C(s_k)x, y_k)$. If dissimilarity measure p defined as (Pulak and Bhabatosh, 2012):

$$p(U, V) = \frac{1}{p} \|U - V\|_p^p, (1 \leq p \leq 2) \tag{8}$$

with $p = 2$, then the estimated HR image satisfies

$$y_k = AH_k C(s_k)x + n_k = B_k(s_k)x + n_k \tag{9}$$

such that:

$$\hat{X} = \operatorname{argmin}_X [\|AH_k C(s_k)x - y_k\|_2^2] \tag{10}$$

A regularization operator $Y(X)$ incorporating prior knowledge was imposed on the estimated HR image X . Then the SR image reconstruction can simply be formulated as:

$$\hat{X} = \text{argmin}_X \{ Y(X) : \|AH_k C(s_k)x - y_k\|_2^2 < \eta \} \quad (11)$$

where η is a scalar constant depending on the noise variance in LR images.

The constrained minimization problem of Equation (11) was reformulated as an unconstrained minimization problem such that:

$$\hat{X} = \text{argmin}_X \{ \frac{1}{2} \|AH_k C(s_k)x - y_k\|_2^2 + \mu Y(X) \} \quad (12)$$

where μ is the regularization parameter that controls the emphasis between the data error term (first term) and the regularization term (second term). To impose regularization to obtain a stable solution of Equation (10), Tikhonov cost function was used to estimate stable de-noised image as presented in Equation (1).

The proposed IBP-MAPsuper resolution reconstruction technique can then be mathematically formulated as:

$$X^{(1)} = X^{(0)} + X_e - X_H^{(0)} \quad (13)$$

where $X^{(0)}$ is the initial interpolated image, X_e is the error correction and $X_H^{(0)}$ is the high frequency estimation given by:

$$X_H^{(0)} = (X^{(0)} * HPF_{BMT}) \quad (14)$$

where $X_H^{(0)}$ is the initial high frequency estimated image, HPF_{BMT} is Bayesian MAP Technique as high pass filter. The estimated HR image after n^{th} iterations is given by:

$$X^{(n+1)} = X^{(n)} + X_e^{(n)} + X_H^{(n)} \quad (15)$$

The estimation of the high frequency is stated as:

$$X_H^{(n)} = X_H^{(0)} - \{ ((y^{(n)} \uparrow S) * HPF_{BMT}) \} \quad (16)$$

where $X_H^{(0)}$ is high frequency component of image $X^{(0)}$ to be obtained from interpolation. So, the final iteration process given in equation (16) can be rewritten as:

$$X^{(n+1)} = X^{(n)} + (y^{(n)} - y) \uparrow S + \{ X_H^{(0)} - \{ ((y^{(n)} \uparrow S) * HPF_{BMT}) \} \} \quad (17)$$

The program flowchart of the developed PBM super resolution reconstruction technique is presented in Figure 3. MAP was used to provide the edge restoration properties of the video frame via the $l1$ norm, non-edge restoration properties and computational efficiencies via the SAR prior, while the $l2$ regularization technique was used to provide the noise-removing. IBP was used to minimize the reconstruction error produced by MAP significantly in an iterative manner and to help realize better and more reliable super-resolved frames in real-time mode. Frames containing facial information from the youtube celebrities video dataset were grabbed using a Video-2-Photo frame grabber application. The low resolution frames were interpolated and projected to the PBM to estimate and produce the desired HR frames.

C. Performance Evaluation Metrics

The numerical comparison between the original and reconstructed HR images was performed using Peak Signal-to-Noise Ratio (PSNR) and Improvement in Signal-to-Noise Ratio (ISNR) as defined below:

- i. Improvement in Signal to Noise Ratio (ISNR) is defined as [29]:

$$ISNR = 10 \log_{10} \frac{\sum_j^M (f_j - g_j)^2}{\sum_j^M (f_j - f_j)^2} \quad (18)$$

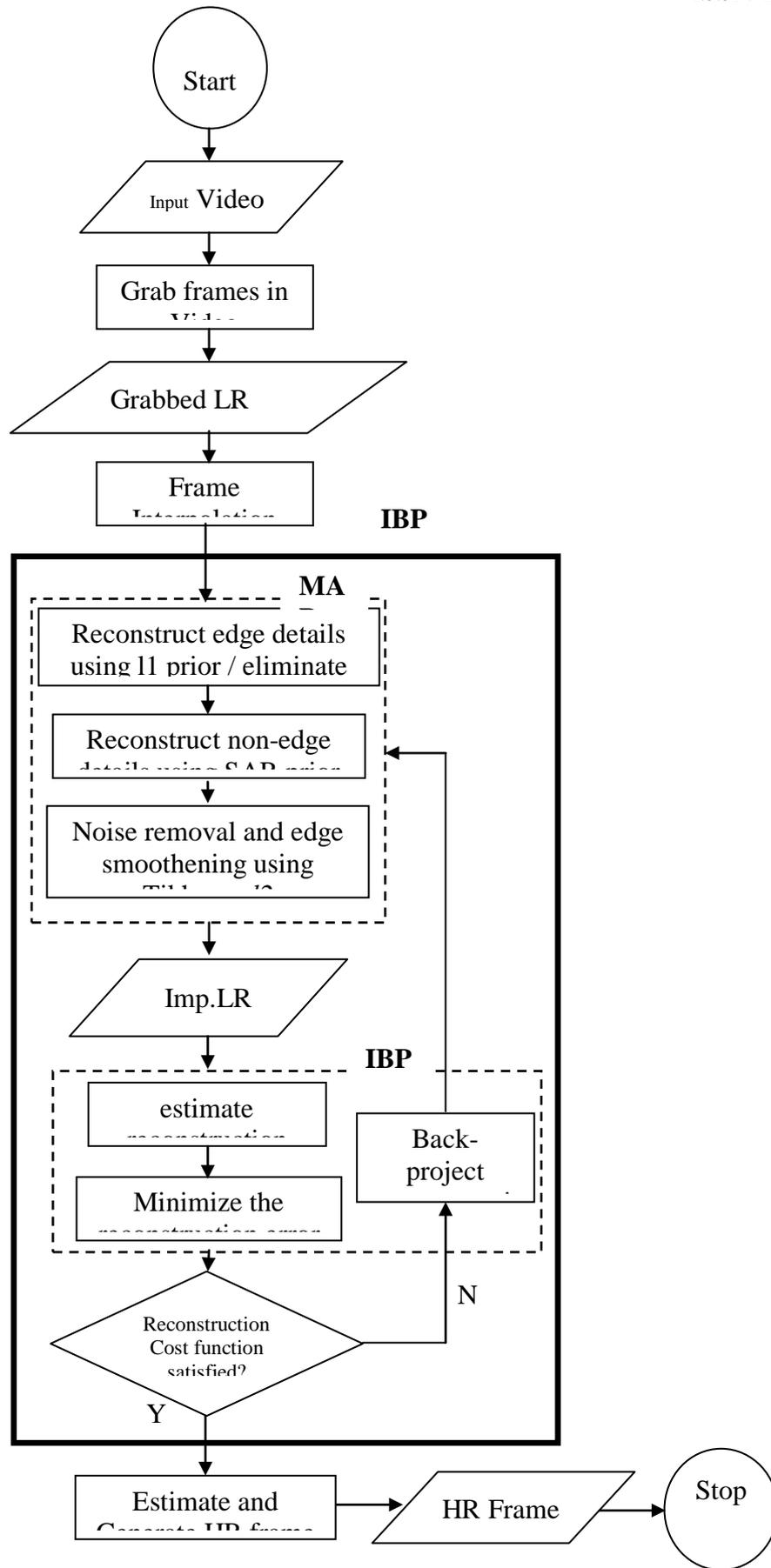


Fig. 3. The Program Flow Chart of the Developed Pixel level Back projected MAP Super Resolution Reconstruction Technique

where M , f , \hat{f} and g denote the total pixel number, the original image, the restored image and the observed degraded image respectively.

ii. Peak Signal-to-Noise Ratio (PSNR) is defined as [25]:

$$\text{PSNR} = 10 \log_{10} \frac{2^x}{\|\hat{x} - x\|^2} \quad (19)$$

where \hat{x} is the estimator of the HR image and x is the true HR image.

IV. Results and discussion

All the super resolution reconstruction techniques were implemented using MATLAB 7.7 (R2008b) on Windows 7 Ultimate 32-bit operating system, AMD Athlon (tm)X2DualCore QL-66 Central Processing Unit with a speed of 2.2GHZ, 2GB Random Access Memory and 320GB hard disk drive. The average of 4 LR images were considered as a blurred HR image. 10 videos were selected from the youtubecelebrities video dataset; 8 frames were grabbed per video and 4 LR images were generated per frame to obtain 32 LR images per video. The LR images were generated on the basis of the hyperparameter λ values. These values range from 0.1 to 0.3. That is, $0.1 < \lambda < 0.3$. The values of λ greater than 0.1 smooth the images while the values below 0.1 make it too noisy. Reconstruction error was obtained as the difference between the LR image and the HR image.

Figures (4-8) present the simulation result for the first frame. However, 32 different frames were super-resolved and the average of the results are presented in figures (9-11); summarized and presented in Table 1. The results obtained indicated that MAP(I1-SAR) produced the highest mean PSNR and ISNR values of 28.577361dB and 10.002076dB respectively than MAP(SAR) and MAP(I1). Since higher PSNR and ISNR translate to better performance, this result indicates that MAP(I1-SAR) performed more optimally than MAP(SAR) and MAP(I1).

On the other hand, IBP produced mean PSNR and ISNR values of 33.270701dB and 11.644750dB respectively which is noticeably higher than those observed in MAP(I1-SAR). In terms of computational complexity, the average number of iterations is 34 for MAP(I1-SAR) and 98 for IBP. However, the developed PBM technique produced the highest mean PSNR and ISNR values of 33.458909dB and 11.710618dB respectively with the least number of iterations. This result indicates that in terms of optimality of performance and computational efficiencies, pixel level PBM is superior to IBP and the MAP variants used as benchmark.

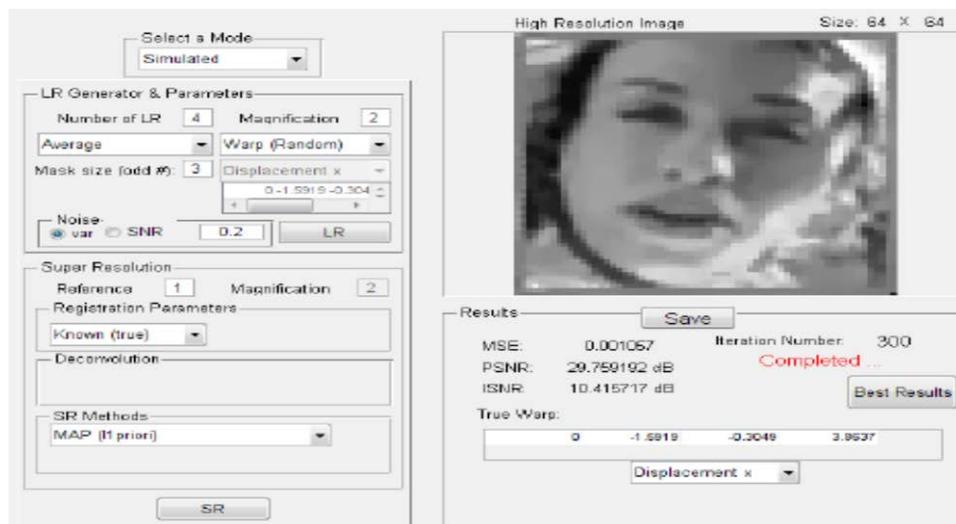


Fig. 4. The MAP (I1) result for frame 1

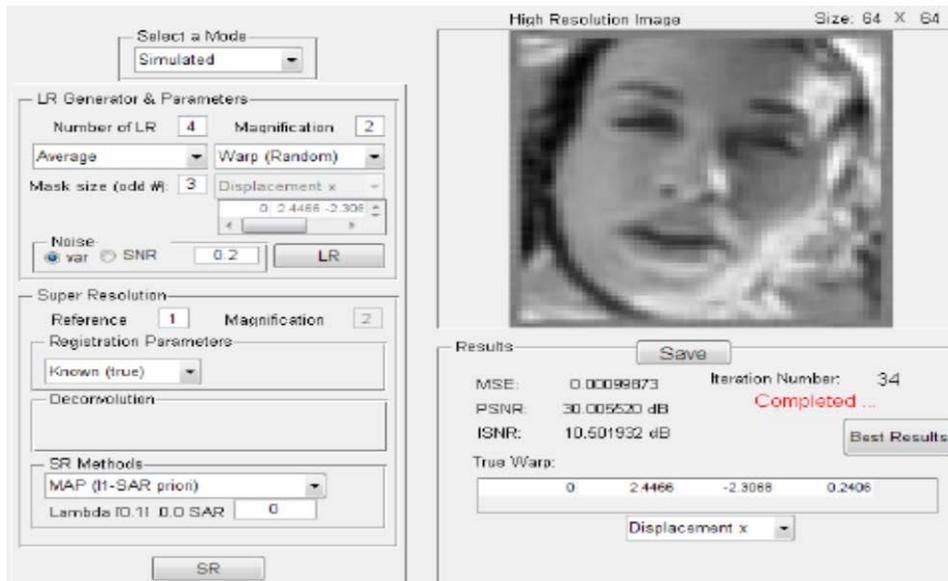


Fig. 6. The MAP (I1-SAR) result for frame 1

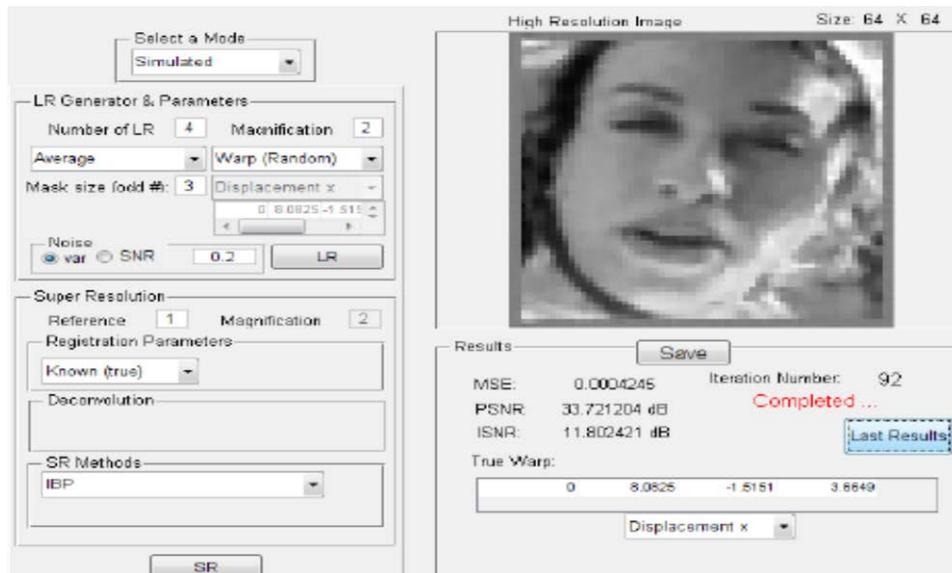


Fig. 7. The IBP result for frame 1



Fig. 8. The PBM result for frame 1

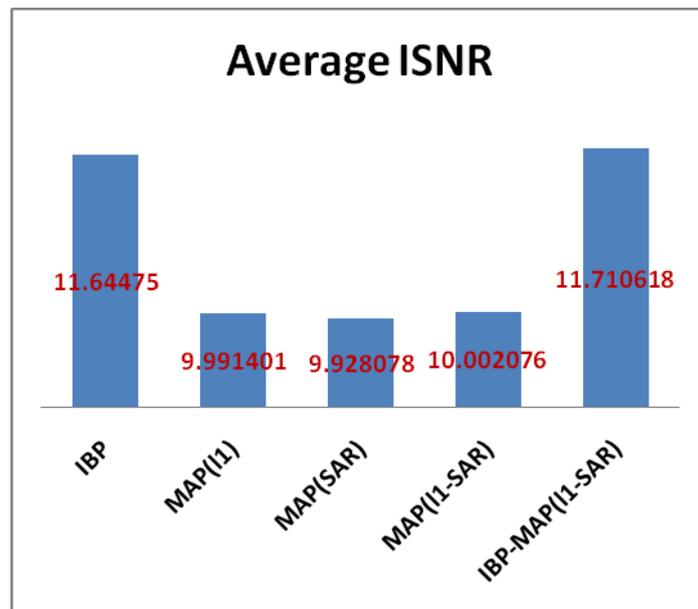


Fig. 9. Average ISNR for the SRR Methods after 32 Consecutive Runs

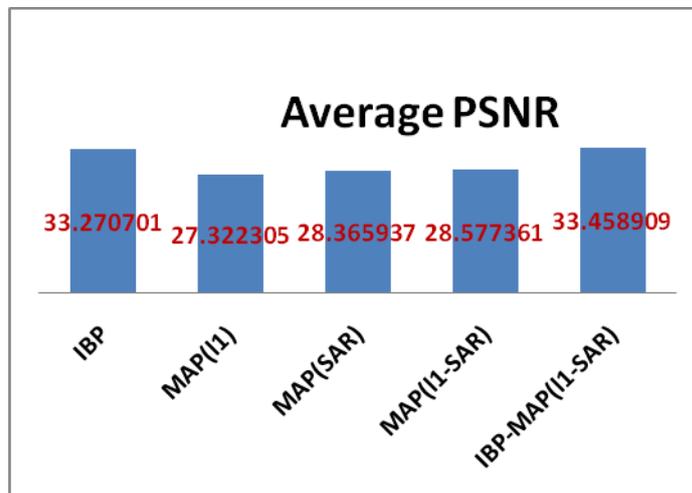


Fig. 10. Average PSNR for the SRR Methods after 32 Consecutive Runs

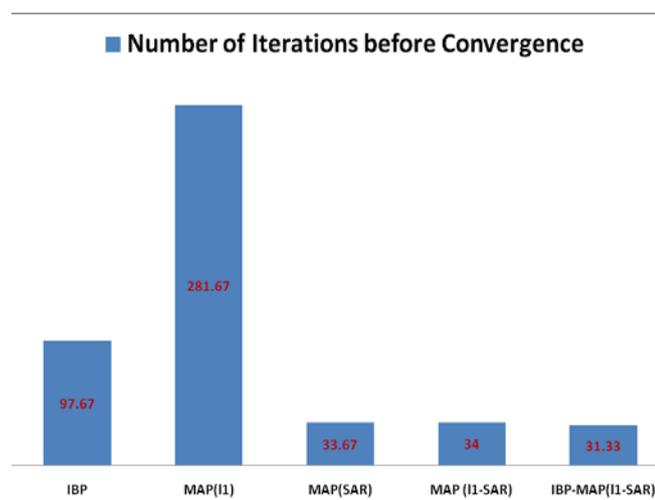


Fig. 11. Average Number of Iterations for the SRR Methods after 32 Consecutive Runs

TABLE I

AVERAGE OF EVALUATION RESULTS OF THE SUPER RESOLUTION RECONSTRUCTION METHODS AFTER 32 RUNS

SRR Method	PSNR (dB)	ISNR (dB)	Nos of Iterations	MSE
IBP	33.270701	11.644750	97.67	0.0004787
MAP(I)	27.322305	9.991401	281.67	0.0026113
MAP (SAR)	28.365937	9.928078	33.67	0.0016091
MAP (I1-SAR)	28.577361	10.002076	34	0.0015520
IBP- MAP (I1-SAR)	33.458909	11.710618	31.33	0.0004541

The results obtained in this paper agree with the assertion of the works of Rujul and Nita [24] and Patel [25] that combining IBP with some other super resolution reconstruction techniques can help to realize better image quality; unfortunately, the authors did not evaluate and report the computational efficiencies of their IBP-oriented hybrid techniques. However, the results from this work negate the report of Patel [25] that IBP is a computationally-efficient technique.

V. Conclusion

In this paper, a computationally-efficient, scalable and more reliable super resolution technique suitable for LR video feeds has been realized. This in turn will help to improve the performance of video-based face recognition systems adopting the solution in real-time mode. Future research work can focus on the development of low-dimensional feature space super resolution reconstruction technique using other sparse methods combined with improved Maximum a Posteriori and investigate its computational efficiencies and accuracy over existing SRR techniques and the PBM technique developed in this paper.

REFERENCES

- [1] E. Alaa, O. Hüseyin and D. Hasan, "Adaptive Fitness Approach - An Application for Video-Based Face Recognition", *New Approaches to Characterization and Recognition of Faces*, InTech, 2012, pp. 153-170.
- [2] S. Sven and B. Wolfgang, "Adjustable Module Isolation for Distributed Computing Infrastructures", *GRID*, 2010, pp. 98-105.
- [3] T. Jason, B. Jeanette, B. Daniel, C. Michael and Z. Heather, "Person Attribute Search for Large-Area Video Surveillance", *Homeland Security Affairs*, 2012, 5(1).
- [4] L. S. Laura, "Local Binary Patterns applied to Face Detection and Recognition", A thesis submitted to the Department of Signal Theory & Communication in partial fulfillment of the requirement for the award of Masters of Science degree in Electrical & Electronics of the Universitat Politecnica De Catalunya, 2010.
- [5] D. Saroj, T. Shilpa, W. Niraj, D. Kailash and K. R. Singh, "Face Detection from Pose Varying Facial Images", *International Journal of Advances in Engineering & Technology*, 2012, 3(2), pp. 286-292.
- [6] E.O.Omidiora, O.A.Fakolujo, R.O.Ayeni and T.M.Ajila. "A Survey of Face Recognition Techniques", *J. of Applied Science, Engineering and Technology*, 2007, 7(1), pp. 57-65.
- [7] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, "Face Recognition, A Literature Survey", *ACM Computing Surveys*, 2003, 35(4), pp. 399-458.
- [8] Motorola, "Video Surveillance Trade-Offs, A Question of Balance, Finding The Right Combination of Image Quality, Frame Rate and Bandwidth" Motorola Solutions, Inc. 1301 E. Algonquin Road, Schaumburg, Illinois 60196 U.S.A, 2012, pp.1-8.
- [9] C. Gabriel, M. Sathish, D. Alexandru, B. Petronel and C. Peter, "Real-Time Video Face Recognition for Embedded Devices", *A Book of Proceedings on New Approaches to Characterization and Recognition of Faces*, InTech, 2012, pp. 153-170.

- [10] J. Vhalos, "Surveillance Society, New High-Tech Cameras Are Watching You", Popular Mechanics, 2008.
- [11] Abbas Bigdeli, Colin Sim, MortezaBiglari-Abhari and Brian Lovell, "Face Detection on Embedded Systems", Retrieved from www.viisage.com, 2006, pp. 1-12.
- [12] C. Jean-François, G. Eric and S. Robert "An Adaptive Classification System for Video-Based Face Recognition" Information Sciences, 2012, 192, pp. 50–70.
- [13] O. Arandjelovic and R. Cipolla, "Achieving Robust Face Recognition from Video by Combining a Weak Photometric Model and a Learnt Generic Face Invariant", Pattern Recognition, <http://dx.doi.org/10.1016/j.patcog.06.024>, 2012.
- [14] W. Huafeng, W. Yunhong and C. Yuan, "Video-based Face Recognition, A Survey", World Academy of Science, Engineering and Technology, 2009, 60, pp. 293-302.
- [15] M.A. Hannan, A. Hussain, A. Mohammed and S. A. Samad, "TPMS Data Analysis for Enhancing Intelligent Vehicle Performance", Journal of Applied Science, 2008, 8, pp. pp. 1926-1931.
- [16] M. Priyadarsini, P. Jasmine and K. Murugesan, "Efficient Face Recognition in Video by Bit Planes Slicing", Journal of Computer Science, 2012, 8 (1), pp. 26-30.
- [17] G. Dmitry, "Video-Based Framework for Face Recognition in Video", [CRVV](http://www.crvv.com), 2005, pp. 330-338.
- [18] C. P. [Christophe](#), G. [Eric](#), S. [Robert](#) and O. G. Dmitry, "Detector Ensembles for Face Recognition in Video Surveillance", [IJCNN](http://www.ijcnn.com), 2012, pp. 1-8.
- [19] P. Anil and S. Jyoti, "Learning Based Single Frame Image Super-resolution Using Fast Discrete Curvelet Coefficients", International Journal of Image Processing (IJIP), 2012, 6 (5), pp. 283-296.
- [20] S. Budi, H. Mochamad and H. P. Mauridhi, "Investigation of Super-Resolution using Phase based Image Matching with Function Fitting", Research Journal of Engineering Sciences, 2012, 1(3), pp. 38-44.
- [21] D. Weisheng, Z. Lei, S. Guangming and W. Xiaolin, "Non-local Back-Projection for Adaptive Image Enlargement", In proceedings of the international conference on image processing, 2009, ICIP, Cairo, Egypt.
- [22] Q. Q. Feng, L. Zhong, H. Z. Li, D. Y. Ying and L. C. Li, "Blind Single-Image Super-Resolution Reconstruction Based on Motion Blur", 10.4028/www.scientific.net/AMR.225-226.895, pp. 895-899, 2011.
- [23] S. Villena, M. Vega, S. D. Babacan, R. Molina and A. K. Katsaggelos, "Bayesian Combination of Sparse and non Sparse Priors in Image Super Resolution", IEEE International Conference on Image Processing, 2012, pp. 1-24.
- [24] R. M. Rujul and D. M. Nita, "Single Image Super-Resolution via Iterative Back Projection Based Canny Edge Detection and a Gabor Filter Prior", International Journal of Soft Computing and Engineering (IJSCE), 2013, 3(1).
- [25] S. A. Patel, "Novel Iterative Back Projection Approach", IOSR Journal of Computer Engineering (IOSR-JCE), 11(1), 2013, pp. 65-69.
- [26] P. Pulak and C. Bhabatosh, "Super Resolution Image Reconstruction Through Bregman Iteration using Morphologic Regularization", IEEE Transactions on Image Processing, 2012, 21(9), pp. 4029-4039.
- [27] J. Yang, J. Wright, T. Huang and Y. Ma, "Image Super-Resolution via Sparse Representation", IEEE Transactions on Image Processing, 2009, 19, pp. 2861–2873.