International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 3 (2015) pp. 5471-5487

http://www.ripublication.com

Detection, Tracking and Identification of Moving Objects in a Video using Super Resolution - A Novel Approach

A.Geetha Devi

Dept., of ECE, PVP Siddhartha Institute of Technology, Vijayawada, India.
geetha.agd@gmail.com

T. Madhu

Swarnandra Institute of Engg& Technology,Narasapuram, India.
tennetimadhu@yahoo.com

K.Lal Kishore
Jawaharlal Nehru Technological University,Ananthapur, India

Abstract

In computer vision, Detection, Tracking and Identification of an intended moving object in a video is a challenging and time consuming task. This paper presents a robust hybrid algorithm to track the required moving object and to reconstruct the image using Super Resolution (SR) algorithms for proper recognition. It uses Corrected Back ground Weighted Histogram (CBWH) method for tracking and multi image super resolution based on Steering Kernel regression based adaptive sharpening filter for restoration and reconstruction. The experimental results demonstrate the effectiveness of the algorithm when compared to other SR algorithms

***Index Terms*—:** Video Tracking, CBWH Histogram, Super resolution, Registration, fusion, Steering Kernel regression based restoration.

1. Introduction

The proliferation of high powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. There are three important steps in video analysis: detection of interesting videos, tracking the objects from frame to frame and analysis of those tracked object to recognize their behaviour

[1]. In the process of identifying the intended object, the tracker has to strive considerably for a long time to scan manually through the low quality video just to locate the object and is based on the tasks of motion based recognition, automated surveillance, video indexing, human computer interaction, traffic monitoring etc.. Consequently, the analysis of the video after tracking becomes a time consuming task.

Tracking objects can be complex due to noise in images, complex object motion, clutter, occlusions in the appearance model of the object to be tracked and loss of information caused by projection of 3D or 2D image [2]. Among the many object tracking algorithms like mean shift algorithms, kernel based algorithms, fast multiple object tracking using particle filters, mean shift algorithm is efficient [3]. Mean shift is a non-parametric feature space analysis technique for locating the maxima of a density function, so called mode seeking algorithm [4]. However the mean shift algorithm is prone to local minima when some of the target features are present in the back ground and target localization cannot be achieved properly. The CBWH (Corrected Background weighted Histogram method) will decrease the weights of prominent background features and improves the target localisation. The application domain includes cluster analysis in computer vision and image processing. This algorithm has been adopted to solve problems mainly in image filtering, image segmentation and object tracking [4].

If the region of interest is enhanced and reconstructed then the task of recognizing the object becomes easy. Although several restoration algorithms are available to eliminate the blur and noise from the detected images, they cannot improve the resolution of the image for better recognition. Super Resolution (SR) Reconstruction is a signal processing technique that develops a high resolution image from the sub pixel information contained by a set of low resolution frames of the video for better visual perception. SR improves the image fidelity as well as automatic recognition rates when dealing with low resolution objects [5]. However, it is a computationally exhaustive process due to the high dimensionality of the reconstruction problem. When a tracker detects and crops the area around the desired object manually, the number and size of the images to be super-resolved can be greatly reduced; as a result there is a net increase in speed.

The paper is organized as follows: Section 2 demonstrates the Video Tracking algorithm, section 3 describes the observation model and back ground of the super resolution algorithm, the methodology of the proposed Super resolution reconstruction is explained in detail in section 4, section 5 presents the results and concluding remarks are discussed in further sections.

2. Intended Object Tracking

Mean shift algorithm is an efficient approach to tracking objects whose appearance is defined by histograms and is one of the best tracking solutions to overcome the robustness to scaling, rotation and partial occlusion [2]. The mean-shift algorithm is useful when the object is represented by its colour density. The algorithm is applied for seeking colour density similarity maxima [3] i.e., the mode of the colour density similarity corresponds to a location of the searched object. Mean-shift is an iterative method that uses the gradient ascent steps to find the local maxima of a given function. The colour density function is a widely used feature within image processing and the mode seeking process comes useful for exploring this feature space.[3] Hence, colour histogram is used to represent the target/object and the pixel weights are formulated in the target candidate model. The main policy of mean shift tracking is to reduce the weight of prominent background features. However, the main drawback of this algorithm is its sensitivity. In order to overcome this, Corrected Back weighted histogram (CBWH) algorithm is used to minimize the sensitivity of mean shift tracking. CBWH further reduces the background interference in the target localisation by reducing the probability of prominent background features both in target model and target candidate model [6]. The object is partitioned into number of fragments and then the target model of each fragment is enhanced by CBWH model. The difference between the fragment and background models lead to weights of background features. CBWH tracks the target very easily and accurately for object recognition [7].

2.1 Target Representation

A feature space has to be chosen to characterise the target. The reference *Target model* is represented by its probability density function *q* in the feature space. The target model is considered to be centred at the spatial location 0 without losing the generality.

In the subsequent frame a target candidate is defined at location *y* and is characterised by its probability density function *p(y)* . Both pdf's have to be estimated from the data. In real time processing of discrete densities m-bin histograms should be used to get low computational cost.

Target Model: $\hat{q} = \{\hat{q}_u\}_{u=1,2.....m}; \sum_{u=1}^{m} \hat{q}_u = 1$ (1)

Target candidate model: $\hat{p}(y) = \{\hat{p}_u(y)\}_{u=1,2,....m}; \sum_{u=1}^{m} \hat{p}_u = 1$ (2)

2.1.1 Target model

Let us denote the pixels in the target region as $\{x^*_i\}$where i=1,2....n, i.e., it contains n pixels. In the target model, the probability of any feature u, which is one of the colour histogram bins is computed by the eq(1),in which $\hat{q}$ is the target model and $\hat{q}_u$ is the probability of u^{th} element of $\hat{q}$ and is given by

$$\hat{q}_u = C \sum_{i=1}^{n} k[\|x_i^*\|^2]\delta[b(x_i^* - u)] \quad (3)$$

$\delta$is the Kronecker delta function,

$b(x_i^*)$ associates the pixel x_i^* to the histogram bin,

k[x] is an isotropickernel profile and constant C is $C = \frac{1}{\sum_{i=1}^{n}(\|x_i^*\|)^2}$

2.1.2 Target candidate model

Let $\{x_i\}$be the normalized pixel locations of the target candidate centred at y in the current frame. Using the same kernel profile k(x) as for the target model with bandwidth h, the histogram representing the target candidate is construed as

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left[\left\|\frac{y-x_i}{n}\right\|^2 \delta(b(x_i)-u)\right] \tag{4}$$

normalized by C_h such that $C_h = \frac{1}{\sum_{i=1}^{n_h} k\left[\left\|\frac{y-x_i}{n}\right\|^2\right]}$ (5)

Since the normalization constant C_h is only dependent on kernel proportions [6], it can be pre calculated when the kernel profile doesn't change in time. This can be achieved only when the size of target candidate remains the same, otherwise when changing h the normalization constant C_h must be recomputed as well.

2.1.3 Mean Shift Algorithm

A key issue in the mean-shift tracking algorithm is the computation of an offset from the current location y to a new location y1 according to the mean-shift iteration equation. The new location y1 is given by

$$y1 = \frac{\sum_{i=1}^{n_h} x_i w_i g\left[\left\|\frac{y-x_i}{h}\right\|^2\right]}{\sum_{i=1}^{n_h} w_i g\left[\left\|\frac{y-x_i}{h}\right\|^2\right]} \tag{6}$$

$$w_i = \sum_{u=1}^{m} \sqrt{\frac{\widehat{q_u}}{\hat{p}_u(y)}}\,[\delta(b(x_i)-u)] \tag{7}$$

where g(x) is the shadow of the kernel profile k(x):g(x)=-k'(x).For the convenience of expression

$$g_i = g\left\|\frac{y-x_i}{h}\right\|^2 \tag{8}$$

Thus, (6) can be re-written as $y1 = \frac{\sum_{i=1}^{n_h} x_i w_i g_i}{\sum_{i=1}^{n_h} w_i g_i}$ (9)

From the mean shift algorithm formula i.e., eq.(9) ,the weights of the points in the target candidate region determines the convergence of the tracking algorithm. The relevant back ground information for the target localisation can only be reduced when the weights of the prominent features are reduced in the mean shift tracking algorithm. This can be achieved through CBWH algorithm and is illustrated as follows.

Let $\hat{o}_u{}_{u=1,\ldots,m}$ (with $\sum_{u=1}^{m} \hat{o}_u = 1$) be the histogram of the background in the feature space and $\hat{o}^*$ be the smallest nonzero entry, where the representation is computed in a region around the target. Let us now denote the weights

$$\left\{ v_u = \min\left(\frac{\overset{\wedge *}{o}}{\overset{\wedge}{o_u}}, 1 \right) \right\}_{u=1,\ldots,m} \tag{10}$$

These weights will transform the target model histogram such that the importance's of those features which have low v_u, i.e., are prominent in the background, will be diminished [6,7]. The background manipulation was already proposed in [7], where the background information was decreased both in the target model and in the target candidate model. The manipulation of just target model histogram that is adopted in this work was proposed in [7]. There is a problem of early termination of the mean-shift algorithm because the shift happens within one pixel after each iteration.

However, in the CBWH algorithm the target model is only transformed but not the target candidate model, and hence the prominent background features are reduced only in the target model, but not in the target candidate model.

Hence the weights

$$w_1'' = \sqrt{\frac{\hat{q}_u'}{\hat{p}_u'(y)}} \tag{11}$$

The denominator component is different from eq., (7) and after simplifying the eq(11) , the weights are

$$w_1'' = \sqrt{\frac{c'}{c}} * \sqrt{V_u'}\ w_i \tag{12}$$

Since $\sqrt{\frac{c'}{c}}$ is constant and is a scaling factor and has no effect on weights. Hence that component can be omitted and the weights can be written as

$$w_1'' = \sqrt{V_u'} \times w_i \tag{13}$$

Equation (13) clearly reflects the relationship between the weight calculated by using the usual target representation (i.e. w_i) and the weight calculated by exploiting the background information (i.e. w''_1). If the colour at a point i in the background region is prominent, then the corresponding value of vu′ is small. Hence in (13) this point's weight is decreased and its relevance for target localisation is reduced. This will then speed up mean shift's convergence towards the salient features of the target. Note that if the background information is not used, then vu′ will be 1 and w''_i will degrade to w_i with the usual target representation. The resolution improvement of the tracked object using CBHW algorithm is realised using Super Resolution algorithm

3. Super Resolution

The SR problem is an ill posed inverse problem in which the aliased and degraded versions of the LR images of the same scene are utilised. The challenge here is to find out an efficient method of SR that enhances the resolution by combining the information in the low resolution video frames.

The low resolution observations can be formulated as

$$y_{c,i} = DHF_{c,i} X_c + N_{c,i} \quad c = \text{R,G,B} \;\; \& \;\; i = 1,2,3 \ldots\ldots\ldots\ldots \text{N}. \tag{14}$$

Where N is the number of low resolution observations made, Xc is the c^{th} colour component of unknown High resolution image, $y_{c,i}$ is the i^{th} Low resolution image of the Xc, D is the down sampling matrix, H is the point spread function of the blur operator, $F_{c,i}$ is the warping matrix and $\mathcal{N}_{c,i}$ is the additive noise.

The estimate of high resolution image can be obtained from the low resolution observations by applying the inverse operations of eq (14). The inverse of the warping operator can be obtained by the registration process in which all the shifted versions of low resolution images are brought in to a single plane with respect to a reference image. The single super resolution image does not contain the entire information of a high resolution image. The details in all the low resolution images have to be fused into a single HR grid using some fusion rule.

The fusion algorithms will decrease the amount of noise and blur but will not remove them completely. Image restoration techniques are employed as a purgative for the noise and blur.

The interpolation process is used as an inverse process of down sampling and bi-cubic interpolation gives smoothed results in many image processing applications and increases the resolution of the image. Bicubic interpolation is the standard used in many commercial image editing software due its better job in obtaining improved resolution than its counterparts.

Depending on the number of input low resolution images SR algorithms are classified as Single image SR and Multiple Image SR algorithms. This paper presents both single image SR and multiple image SR algorithms for the reconstruction of the tracked image to compare with the proposed algorithm. The single image SR algorithms utilize dictionary based algorithms [8,9] and the multiple image SR reconstruction consists of Registration, Fusion, Interpolation and Restoration stages.

4. Proposed algorithm

The proposed SR algorithm is a multi image SR technique as shown in the Fig.1, The region of interest that has been tracked using the CBWH tracking algorithm in the consequent frames has been taken as the LR input to the Super Resolution algorithm for reconstruction. All these frames are registered using the Automatic Feature Based Registration Using Scale Invariant Feature Transform **(**SIFT). The Registration algorithm used to bring all the LR frames into a common geometrical plane is explained as follows:

4.1 Automatic Feature Based Registration Using Scale Invariant Feature Transform (SIFT)

Registration is the process of alignment of all the shifted versions of low resolution images into a single geometrical plane with respect to a reference image. Feature based registration[10] provides better results when compared to other types of registration processes[11,12,13] and comprises of feature detection, feature matching, optimum transformation and up-sampling and provides better results in many applications. Features of the image are the distinct and prominent objects like edges, lines and contours which can be detected either manually or automatically. These points are called as the control points and their location and scale has to be determined by a detailed model. SIFT (scale Invariant Feature Transform) model is used for the automatic registration [14].

In SIFT algorithm the control points are called as SIFT keys and are obtained using the Difference of the Gaussian (DoG) by comparing a pixel with its 26 neighbours at the current three adjoining scales and based on the image gradient directions each key point location is given one or more orientations. The feature matching establishes correspondence between the detected features of the image. The regular approach is to build local descriptors around the feature point and then match the descriptors. This is a very important step since the amount of accuracy in the correct match's identification decides the precision of the transformation in the next step. Euclidian distance matching, invariant moment and nearest neighbour based matching are the usual methods of feature matching. RANdom SAmpling Consensus (RANSAC) is a strong feature estimator and is proposed in the year 1981 by Fischeler and Bolles [9]. It classifies the matching features into inliers and outliers. Inliers are the features that hold on to the model while the outliers won't. The RANSAC algorithm starts by randomly selecting the set of corresponding points.

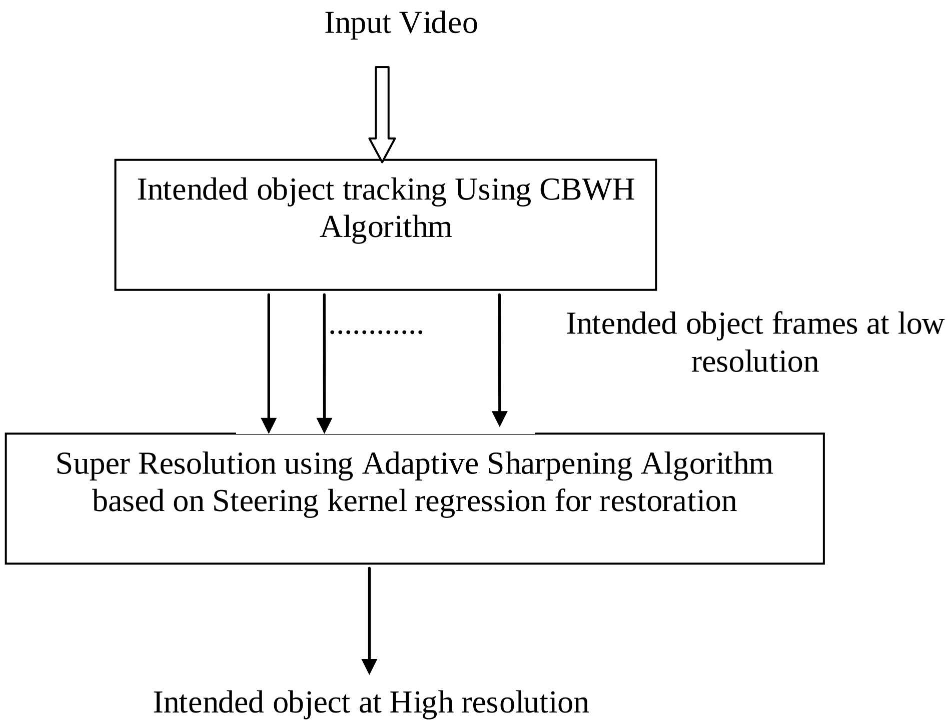

Fig.1: Block diagram representation of the Proposed Algorithm

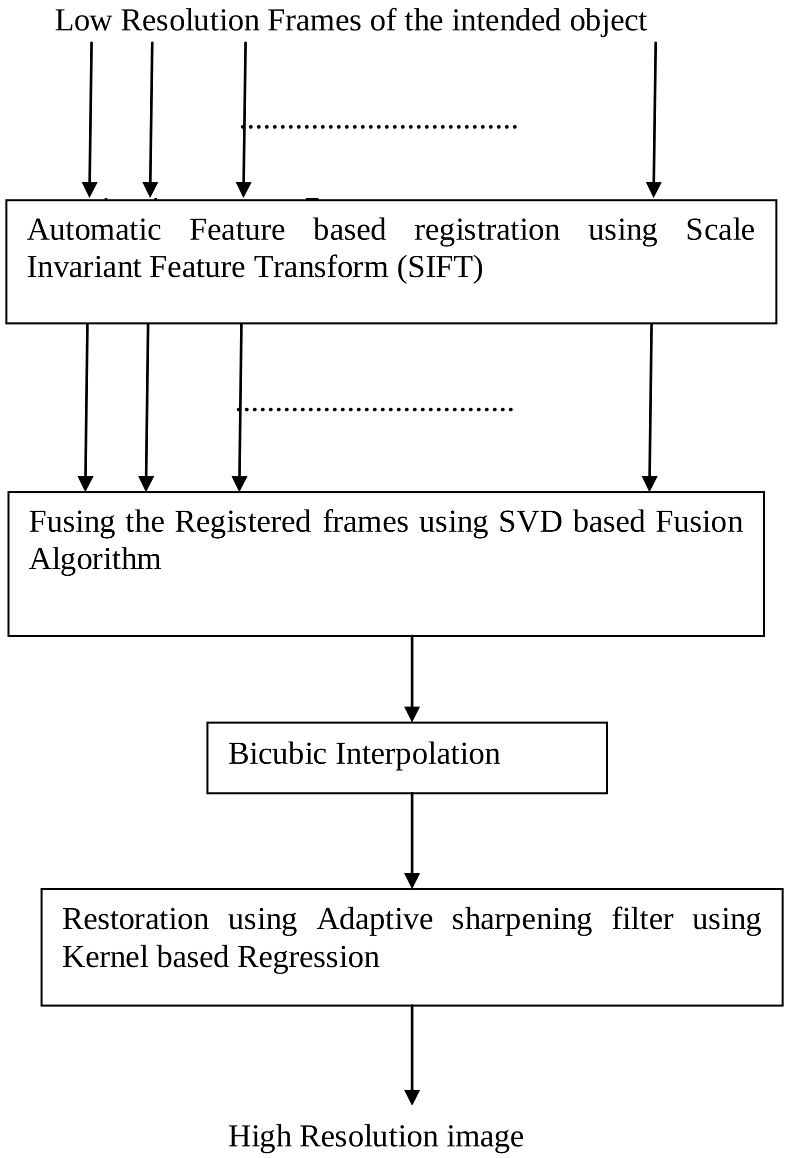

Fig.2: Steps followed in Super Resolution Algorithm

For each possible set of four key points in the reference image and their corresponding match in the target image, a mapping transform is found. Then a transformation matrix is estimated. The symmetric transfer error for every matching point and the inliers that are less than the threshold value are counted. Then the same procedure is applied to the rest of the key points in the reference image and the spatial coordinates of transformed key points are compared with the coordinates of the respective key points in the target image. This allows a number of key point pairs that fit the model to be identified within a certain tolerance. The model that supports the maximum number of key point pairs (consensus set) within a transform model is considered as optimal. After finding the optimal value, the model will transform the target image into the reference image, so that the corresponding points in both the images are spatially coincident. After registering the images the sub pixel information in the LR images has to be fused into a single image to obtain more details of the High Resolution image.

4.2 Singular Value Decomposition (SVD) Fusion

Image fusion is the process of integrating the information contained in all the LR observations into a single image. The Singular Value Decomposition (SVD) fusion gives better results in applications like signal processing, pattern recognition and data compression applications [15,16].

The SVD of any matrix L of dimension m×n is represented by

$$L=USV^T \tag{15}$$

Where the matrices U and V are orthogonal to each other.

The columns of the m×n matrix U are the eigen values of the LL^T and is known as left singular vector matrix and the columns of the m×n matrix V^T are the eigen values of the L^TL and is called the right singular vector matrix. The diagonal elements of the n×n matrix S are the singular values of the matrix L. It represents the intensity information of *'L'*. The gray scale representation of any image is a two dimensional matrix which can be decomposed in to SVD. Here, the highest singular value has greatest amount of input information in it and the change of highest value of SVD lies at the upper left corner of the S matrix.

The SVD fusion of the two images L_1 and L_2 are represented respectively as

$$L_1=U_1S_1V_1^T \tag{16}$$

$$L_2=U_2S_2V_2^T \tag{17}$$

For the colour images the decomposition is performed in each colour plane separately. Let the maximum values of S_1 and S_2 are $\beta1_{max}$ and $\beta2_{max}$ respectively. Then if $\beta1_{max} > \beta2_{max}$ then S_1 is used in the reconstruction of the fused image otherwise S_2 is used.

$$L_{fused}=U_2\, S_{max}V_2^T \tag{18}$$

Where $S_{max} = S_{1;}$ $if\, \beta1_{max} > \beta2_{max}$
$=S_{2;}$ $if\, \beta2_{max} > \beta1_{max}$

When $\beta2_{max} > \beta1_{max}$ the reference image will not come into picture at all. But in most of the cases $\beta1_{max} > \beta2_{max}$, which forms a fused image of original and reference images.

The fused image L_{fused} is passed through the interpolation step. When more than two images are available for registration the approach can be extended to multiple images.

4.3 Bicubic Interpolation

The resolution of the image is improved by preserving the finer details of the fused LR images during interpolation. Bicubic Interpolation determines the grey level value from the weighted average of the 16 closest pixels to the specified input coordinates, and assigns that value to the output coordinates. The image is slightly sharper than that produced by Bilinear Interpolation, and it does not have the disjointed appearance produced by Nearest Neighbour Interpolation.

The intensity of the pixel is computed by considering its sixteen nearest neighbours as

$$x(p,q) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij} p^i q^j \tag{19}$$

Where the sixteen coefficients a_{ij} are determined from sixteen neighbours. Solving the sixteen equations provides a surface *x(p, q)* on unit square which is continuous and with continuous derivatives. Bicubic interpolation on a random sized regular grid can then be achieved by patching all such cubic surfaces, making sure that the derivatives match on the boundaries. If the derivatives are not known then they are typically approximated from the function values at points neighbouring the corners of unit square [17].

In the proposed algorithm the restoration algorithm is a steering kernel regression[18,19,20] based sharpening filter which takes the interpolated image as its input and is explained as follows:

The observation model of the interpolated image with noise and blur is given by

]

$$y=hX+n \tag{20}$$

where *X* is the HR image and *h* is the degradation due to blur, warping and down sampling.

In order to perform the sharpening and denoising operations on the image simultaneously, the sharpening kernel *S* should satisfy $Sh \approx I$, Unity matrix since *S* has to invert *h* globally.

The estimated output image after sharpening is given by

$$\hat{X}=S\,[hX+n_k]= ShX+S\,n_k \tag{21}$$

The second term amplifies the noise. Hence to avoid the amplification of noise the sharpening kernel should be adaptive to the local image characteristics i.e., the filter should concentrate more on noise reduction in the areas where the effect of blur is not felt. In the places where there is more edge information, the filter should sharpen the image in the edge direction only. Such type of adaptive sharpening filter is described as follows.

4.4 Steering Kernel Construction

Assuming a pixel of interest x_i, its Steering kernel (SK) is mathematically represented as

$$sk(x_l - x_i) = \sqrt{|c_l|}\exp\{-(x_l - x_i)^T c_l (x_l - x_i)\} \quad (22)$$

Where x_l is the given location inside the SK window centred at x_i and c_l is the covariance matrix estimated within SK.

A separate covariance matrix is also estimated and used at each pixel location as it results in a far richer set of shapes for the resulting kernel weights than would otherwise be expected as a Gaussian kernel. In the flat region of an image the SK is basically isotropic indicating that there is no strong directional structure in that area. In the edge region the shape of SK illustrates the edge outline, and the kernel values mainly indicate the pixel intensity similarity with respect to the pixel of interest. In the region that contains small scale image details, the related SK also shrinks to a smaller region and it is observed that the *Sk* s are basically robust to high levels of noise. The kernel values are computed as follows: The gradient matrix D for every window located at each pixel location is calculated and the dominant direction V_1 and its perpendicular direction V_2 with in a window can be estimated from the gradient matrix using the singular value decomposition (SVD) as

$$D=U\ A\ V^T = U\begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}\begin{bmatrix} v_1 & v_2 \end{bmatrix}^T \quad (23)$$

In eq.(22) S_1 and S_2 are the singular values which represents the energy in the directions V_1 and V_2 respectively and $S_1 \geq S_2 \geq 0$.

The covariance matrix Cl can then be estimated using the following formula

$$c_l = \alpha(\lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T) \quad (24)$$

Where α is scaling parameter and λ_1 is elongation parameter and is given by

$$\lambda_1 = \frac{s_1 + \xi'}{s_2 + \xi''},\ \lambda_2 = \frac{1}{\lambda_1}$$

$$\text{and} \quad \alpha = \left(\frac{s_1 s_2 + \xi''}{M}\right)^k \quad (25)$$

In the eq.(23) ξ' and ξ'' are the regularization parameters that dampen the noise effect and restrict elongation and scaling parameters to become zero. *M* is the number of samples of the analysis window considered and the scalar k controls the local kernel that is affected by local window.

Soon after the steering kernel *Sk* is obtained, The sharpening kernel is constructed by using

$$S=Sk+qL*Sk \quad (26)$$

Where * represents the convolution operation and L is the Laplacian operator. The value of q determines the degree of sharpening and the restored image is computed as a local weighted average with the steering kernel *Sk* as follows

$$\widehat{f}(x_i) = \frac{\sum_{x \in wl} Sk(x - x_i) g(x)}{\sum_{x \in wl} Sk(x - x_i)} \tag{27}$$

Where *g(x)* is measured on the LR image. The sharpening process can be represented in the vector form as

$$\widehat{f} = D_{sk}^{-1} S g = D_{sk}^{-1} (Sk + QSkL) g \tag{28}$$

In the above eq., *S* is the sharpening matrix, *Sk* is the steering kernel matrix and D_{sk} is the diagonal matrix of the row sums of the matrix *S* which acts as a normalization factor. *Q=diag(q1,q2,q3……)* assigns to each pixel the corresponding sharpening parameter.

The image sharpness metric in the presence of noise is implemented by singular values of the gradient matrix, which have been already calculated in the steering kernel construction.

The local metric *Q* for the pixel located at xi is defined as

$$Q = s_1 \frac{s_1 - s_2}{s_1 + s_2} \tag{29}$$

The local image patch has more blur component as Q value is less and to avoid over shoots in already sharp edges, the parameter q has to be set.

$$q = \begin{cases} \beta & if \quad Q < T_1 \\ \beta(Q - T_2)/(T_1 - T_2) & if\ T_2 < Q < T_1 \\ 0 & if \quad Q \geq T_2 \end{cases} \tag{30}$$

5. Results& Discussion

The intended object that is traced in the video has to be reconstructed by using the proposed SR algorithm. Since the proposed method is a multi image SR, the LR images from four consecutive frames are applied as an input to the SR reconstruction algorithm. The LR tracked frames for four videos are shown in the Fig.3. The Single image Super resolution using sparsity and multi image SVD based SR using blind deconvolution algorithms are utilised for the comparison of the proposed algorithms for different input databases. The difference between the proposed algorithm and the other algorithms for the first two databases vipman and vipface are demonstrated in Fig 4. Fig.5. shows the HR images reconstructed for the viptrain video and Fig.6. gives the reconstructed images for the tracked portion of the book_case_small video. The comparison of these algorithms for different input videos are accomplished using mean, standard deviation and entropy and are listed in the tables1,2&3. From the results it is clear that the smaller regions in the detected image are clearly reconstructed in the proposed algorithm when compared to other SR algorithms. In the vipman database the facial information is more clear in the proposed algorithm. In the vipface database the area around the eyes is restored very clearly when compared to the other SR algorithms and also the other facial information is reconstructed in a better manner and provides smoothed output. The output HR image in the vip train database is reconstructed such that it has prompt edges when compared to the HR images in the other techniques. The book names in the book_case database are legible clearly and the edges are connected to each other. From the results it is shown that the proposed algorithm is providing better visual quality by removing the blur and noise and preserving the colour information.

Fig3: The four consecutive Low Resolution tracked video frames of (a)Vipman (b) vipface (c) viptrain (d) book_case small

Table1: Comparison of reconstructed data using Mean

Video input ↓	Mean		
	Single image SR	SVD based SR using blind deconvolution	Proposed Algorithm
Vipman	121.482	123.517	126.462
Vipface	86.42	85.37	86.77
Viptrain	114.2021	111.278	115.79
Book_case_small	105.55	109.94	111.59

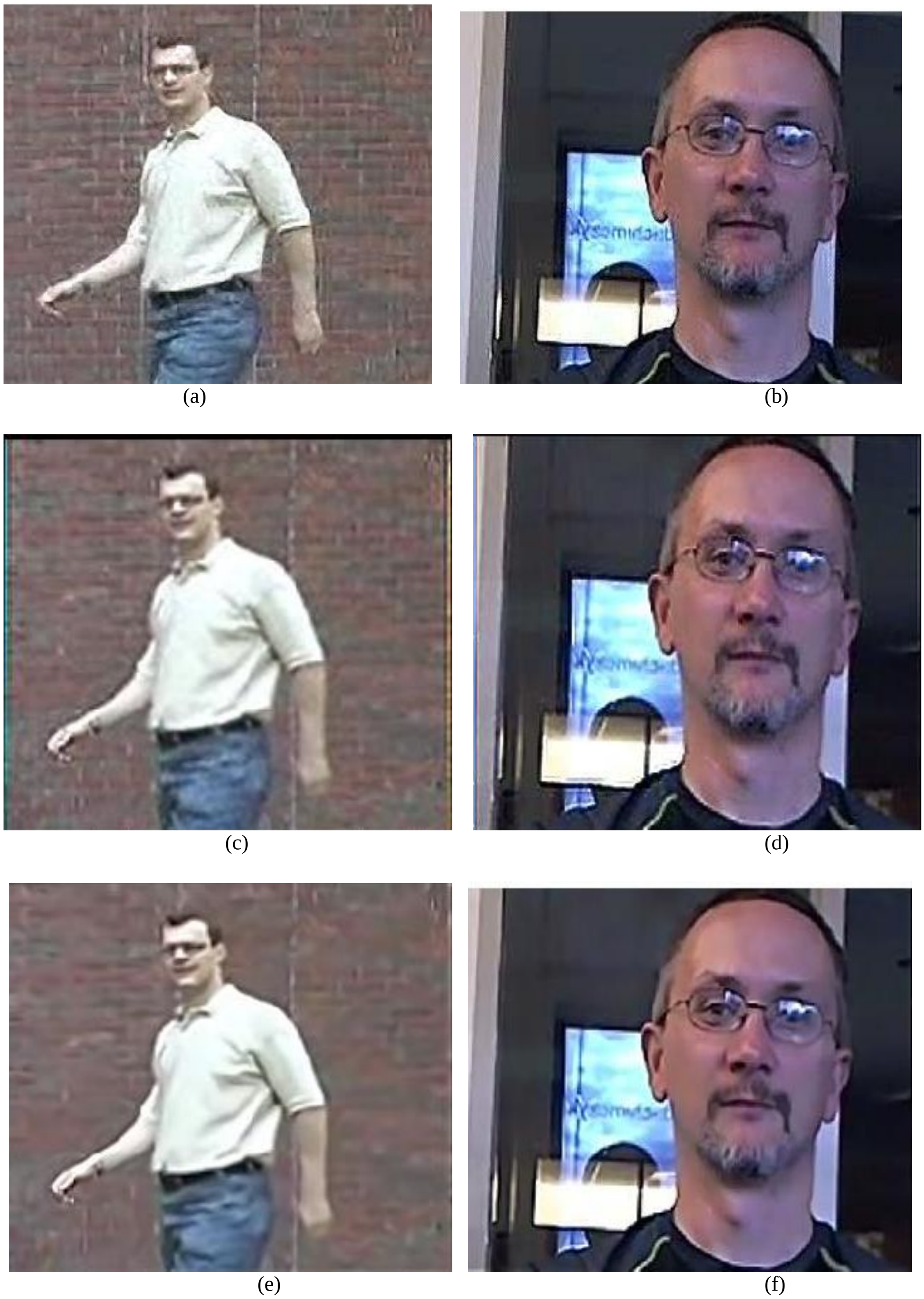

Fig: 4. Reconstructed images of Vipman and vipface tracked video (a) &(b) Single image SR reconstruction using sparse (c)&(d) SVD based SR using blind deconvolution (e)&(f) Reconstructed image using Proposed algorithm

Fig:5. Reconstructed Images from the viptrain tracked video using (a) Single image SR reconstruction using sparse (b) SVD based SR using blind deconvolution (c)Reconstructed image using Proposed algorithm

Fig:6. Reconstructed Images from the Book_case1_small tracked video using a) Single image SR reconstruction using sparse b) SVD based SR using blind deconvolution c) Reconstructed image using Proposed algorithm

Table2: Comparison of reconstructed data using Standard Deviation

Video input ↓	Standard deviation		
	Single image SR	SVD based SR using blind deconvolution	Proposed Algorithm
Vipman	48.52	49.74	52.35
Vipface	61.477	63.2928	62.6641
Vip train	73.28	77.79	79.08
Book_case_small	72.46	77.72	78.63

Table3: Comparison of reconstructed data using Entropy

Video input ↓	Entropy		
	Single image SR	SVD based SR using blind deconvolution	Proposed Algorithm
Vipman	6.73	6.71	6.89
Vipface	7.46	7.37	7.57
Vip train	7.46	7.54	7.58
Book_case_small	7.54	7.52	7.74

Conclusions

This paper proposed a novel approach for object recognition. The CBWH algorithm which can detect the target accurately and easily is integrated with a multi image Super Resolution algorithm to simultaneously improve the resolution and reconstruct the detected target under noisy environment. The SR algorithm is applied with an adaptive Steering kernel regression based edge sharpening filter to remove the noise from the image while preserving the edge information and improving the colour information. The future work will be to improve the SR algorithm for enhanced recognition of the intended object.

References

[1] Yilmaz, A., Javed, O., Shah, M., 2006, “Object tracking: a survey”, *ACM Computing Surveys*, Dec. 38, (4), Article 13

[2] Anand Singh Jalal, Vrijendra Singh, 2012 “The State-of-the-Art in Visual Object Tracking”, *Informatica 36* , ,pp:227-248.

[3] Cheng, Y.: 1995, “Mean shift, mode seeking, and clustering”, *IEEE Trans. Pattern Anal. Mach. Intell.*, , 17, (8), pp. 790–799

[4] Comaniciu, D., Meer, P., 2002 “Mean shift: a robust approach toward feature space analysis”, IEEE Trans. Pattern Anal. 24, pp:603–619

[5] Sung Cheol Park, Min Kyu Park, Moon Gi Kang , May 2003, “Super Resolution Image Reconstruction: A Technical Overview”, IEEE Signal processing magazine, vol 20 no.3 pp: 21-36.
[6] Comaniciu, D., Ramesh, V., Meer, P.,2003, “Kernel-based object tracking”, IEEE Trans. Pattern Anal. Mach. Intell, 25, (2), pp. 564–577.
[7] J. Ning,L. Zhang, D. Zhang, C. Wu, 2012, “Robust mean-shift tracking with corrected background-weighted histogram”, IET Computer Vision, Vol. 6, Issue.1, pp. 62–69
[8] M. Aharon, M. Elad, and A. Bruckstein., 2006, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation”, IEEE Transactions on Signal Processing, Vol 54, No. 11, Novermber
[9] M. Elad and M. Aharon., 2006, “Image denoising via sparse and redundant representations over learned dictionaries”, IEEE Trans.on Image Processing, Vol.15, No. 12.
[10] Arun K.S.,Sarath K.S., 2010, “An Automatic Feature Based Registration Algorithm for Medical Images”, International Conference on Advances in Recent Technologies in Communication and Computing, IEEE Computer Society
[11] Brown Gottesfeld L., 1992, “Survey of Image Registration techniques”, ACM Computing Surveys,Vol.24,Issue 4, pp. 325-376
[12] Barbara Zitova, Jan Flusser, 2003, “Image Registration Methods: A survey”, Image and Vision Computing Vol.21, pp.977-1000
[13] J.B. Antoine Maintz and Max A. Vierger, 1/98, “A Survey of Medical Image Registration, Medical Image Analysis”, volume 2. number1, pp 1-37.
[14] Z. Yuan. P. Yan, and S. Li, July 2008, “Super resolution based on scale invariant feature transform,”, Proc. Int. Conf. Audio, Language and Image Processing, ICALIP 2008, pp:1550-1554.
[15] Haidawati Nasir,Vladimir stankovic, Stephen Marshall, 2012, “ Singular Value decomposition based Fusion for super resolution Image Reconstruction”, Signal Processing; Image Communication 27, pp:180-191
[16] A. Geetha Devi, T. Madhu and K. Lal Kishore, 2014, “An Improved Super Resolution Image Reconstruction using SVD based Fusion and Blind Deconvolution techniques”, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.7, No.1 pp.283-298.
[17] Rafael C. Gonzalez, R. E. Woods, Stevan L.Eddins, 2003, “Digital Image Processing” Pearson Education
[18] B. Zhang and J. P. Allebach, May 2008, “Adaptive bilateral filter for sharpness enhancement and noise removal”, IEEE Trans. on Image Processing, 17(5):664–678.
[19] H. Takeda, S. Farsiu, and P. Milanfar, February 2007, “Kernel regression for image processing and reconstruction”, IEEE Transactions on Image Processing, 16(2), pp:349–366
[20] Xiang Zhu and Peyman Milanfar, 2011 “Restoration for Weakly Blurred and Strongly Noisy Images”, Applications of Computer Vision (WACV), 2011 IEEE Workshop on, 5-7 Jan, pp:103-109.