USING ANAGLYPH 3D TECHNOLOGY FOR VIDEO PRODUCTION WITH HIGH RESOLUTION BASED ON SUPER-RESOLUTION[+]

Zina Abdul Lateef [*]

المستخلص:

الغرض من البحث هو تحسين الصور الفيديويه ذات الدقة الواطئة والصغيرة الحجم بعد تحسينها وتحويلها الى صور فيديوية ثلاثية الابعاد كما هي المعروضة في التلفاز والسينما، بالاعتماد على خزين من البيانات المتراكمة لأفلام مماثلة لغرض الوصول إلى دقة عالية في تحسين اللقطة المشوشة الحالية. حيث تؤخذ مجموعة من لقطات الفيديو ويتم مقارنة بيانات اللقطة المشوشة مع مجموعة كبيرة من البيانات الخاصة بلقطات واضحة مماثلة ذات وضوح عالي لغرض الوصول إلى أقرب حالة تطابق مع اللقطة المشوشة والتي سيتم استبدالها باللقطة الواضحة المقاربة لها، على أن يتم قبل ذلك التأكد من الانسيابية في اللقطات من خلال مقارنة اللقطة السابقة واللاحقة مع اللقطات المناظرة لها في الفلم الواضح، وكذلك بالاستفادة من حركة الجسم الموجود في اللقطة المشوشة.هذه هي التقنية المتبعة في أسلوب مقارنة اللقطات المشوشة مع اللقطات المماثلة الواضحة التي تلتقطها العدسة المحدبة وتعرف بـــ(POCS) لغرض التحديد الدقيق لقيمة جميع النقاط المكونة للصورة.وقد أظهرت النتائج الابحاث) بأن تحسن في نوعية الفيديو يمكن تحقيقه وتحويل الصورة ذات البعد الثاني الى البعد الثالث بالاعتماد على تقنية الـــ Anaglyph 3D.

ان التقنيات المرئية والصور المرئية ثلاثية الابعاد تعتبر من التنقيات المهمة في الوقت الحاضر وخصوصاً لما لها من استخدامات في العرض السينمائي والتلفازي. وفي بحثنا هذا، تم اخذ لقطات صورية فيديوية وتم تكبيرها ومن ثم استخدام تقنية anaglyph الثلاثية الابعاد لغرض تحويل اللقطة الفيديوية المحدده بعد استخلاصها من الفلم الفيديوي وتحويلها ثلاثية الابعاد بعد استخدام النظارات الخاصة بها.

استخدام تقنية anaglyph ثلاثية الابعاد لتحسين الصور الفيديوية بدقة عالية بالاعتماد على الدقة الفائقة

زينه عبداللطيف

Abstract:

Our approach based on enhancing the resolution of video and converted to an anaglyph 3D as used in TV and other application, the image which captured from sequence of video images by perceptually adding plausible high frequency information. By introducing an appropriate prior distribution over such data set we can ensure consistency of static image regions across successive frames of the video, and also take account of object motion. A key concept is the use of the previously enhanced frame to

[+]**Received on 121/2012 , Accepted on 29/4/2013 .**
[*]**Assistant Lecturer / Musayyib Technical Institute**

provide part of the training set for super-resolution enhancement of the current frame and that by adopted the local variance of reference frame as the foundation to establish the operator for Projection Onto Convex Set (POCS) and local threshold value for pixel-repair. The results show that an improvement in images sequence quality can be achieved. Other phase of this approach is take a frame that enhanced and will discuss principles of stereoscopy then moved to the 3D image by converting an image (image taken from process of super-resolution). Also, the results will use one of 3D technology which is Anaglyph and it's known as two colors (red and cyan) with many different images with consideration of different details found in image.

3D imaging and visualization technologies are commonly considered to be the most novel, recent, and advanced form of visual content delivery, 3D photography, cinema, and TV actually have a long history. In this proposed work, implement the last process of the frame which took from a video and process by Super Resolution technology, and processed by Anaglyph 3D technology to convert the processed frame (image) to 3D vision which be wear anaglyph class for this kind of vision.

Keywords: super-resolution, LR, HR, deblurring, Anaglyph 3D, Stereoscopy

Introduction:

Video sequences usually contain a large overlap between successive frames, and regions in the scene are sampled in several images. This multiple sampling can sometime be used to achieve images with a higher spatial resolution. The process of reconstructing a high resolution image from several images covering the same region in the world is called Super Resolution. Additional tasks may include the reconstruction of a high resolution video sequence[1].

Due to the different positions of our eyes, while observing the environment, two slightly different 2D images of the 3D scene fall onto the retina of each eye. The human visual system is based on (a) capturing light via our pupils and (b) processing it via the lens and the variable size of the pupil before receiving a focused color image on the retina. Stereoscopy is based on the capture of those two slightly different 2D images that mimic the images naturally obtained by two eyes at slightly different positions, followed by the simultaneous delivery of each image to the corresponding eye. Different modes of separation techniques have been used. The capture is usually accomplished by two parallel cameras. Ideally the two cameras should physically be the same, and the alignment should be perfect. Usually, special glasses are needed for separating the two images. Early versions of these glasses were based on anaglyphs, which are two intensity images, each having a different color whose color spectra do not overlap. Each of the two images is printed (photography), or projected (cinema), or electronically displayed (TV) by overlapping them properly. Eyewear consists of different color filters on each eye that filter, and thus separate, these overlapped images. The color filters should match the spectrum of the displayed images so that each eye receives only its single 2D intensity image. This technique is still used in printed stereoscopic photography. Due to the restrictions on the color, full-color anaglyph-based stereograms are impossible. However, more sophisticated means of filtering are adopted for cinema and TV [2].

In proposed design is based on applying high super resolution method to frames which has been extracted from video, then applied Anaglyph technique to high super resolution frames. Finally, the results were a 3D image (anaglyph) for different size and resolution of video, besides that, the MSE and PSNR has been calculated.

Anaglyphs can be used in many applications including video games, theatrical films, DVDs and in science. In science the depth perception has many uses. It can reveal features by improving the visualization quality of geo- morphological maps [3].

Super-Resolution Reconstruction:

The reconstruction process of the reconstruction-based super-resolution can be formulated as an inverse problem that estimates the source of the high resolution image from observed low-resolution images assuming an image generative model. Several image generative models have been proposed for SR reconstruction. A widely used image generative model is[3]

$$y_i = D\,H_i F_i\,x + v_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where $y_i$is the vectorized$i$-th observed image, x is the vectorized high resolution image, $F_i$is a matrix representing the motion of the i-th image, $H_i$is a matrix representing the point-spread function (PSF) of the i-th image, D is a matrix representing the down sampling, and vi represents the noise of the i-th image.

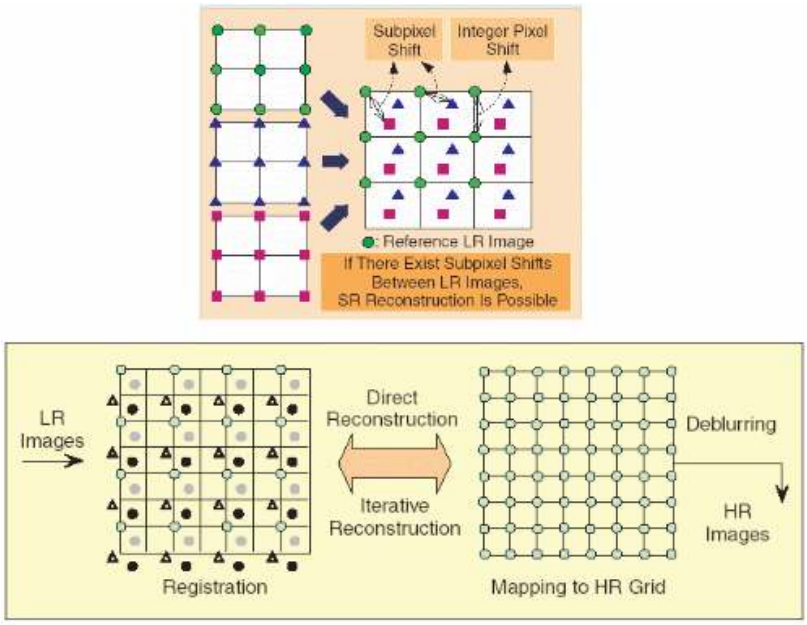


Figure (1) Super Resolution Reconstruction

A maximum a posteriori (MAP) estimation is the popular solution of the inverse problem associated with the generative model in Eq. (2). The MAP estimation is performed by minimizing the cost function:

$$E = \sum_{i=1}^{N} \|y_i - DH_iF_ix\|_2^2 + \alpha\lambda(x) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where $\lambda(x)$ is a constraining function derived from the prior distribution of the high-resolution image, α is a hyper-parameter, and N is the number of observed images. The cost function is the negative logarithm of the posterior distribution of the high-resolution image when the observed low-resolution images are given, wherein the noise model is assumed to be independent Gaussian distribution[4].

Robust SR Reconstruction:

First, by show the pixel-wise cost function to express a general robust cost function. The pixel-wise generative model can be derived from Eq. (3) as[3]

$$y_{i,j} = b_{i,j}^{T} . x + v_{i,j} \ldots\ldots (3)$$

where *yi,j*is the *j*-th pixel value of the *i*-th observed image, *vi,j*represents the noise of the *j*-th pixel value of the *i*-th observed image, and $b_{i,1}^{T}$ is defined as

$$DH_i F_i = \begin{pmatrix} b_{i,1}^{T} \\ b_{i,2}^{T} \\ \ldots \end{pmatrix} \ldots\ldots (4)$$

The cost function of the MAP estimation in Eq. 5 can also be rewritten as a pixel-wise expression,

$$E = \sum_{i=1}^{N} \sum_{j=1}^{M_i} [y_{i,j} - b_{i,j}^{T} . x]^2 + \alpha\lambda(x) \ldots\ldots (5)$$

where $M_i$is the pixel of the *i*-th observed image.
The occlusion and/or the registration error are not involved in the pixelwise generative model. Therefore, we must reject the occluded pixels and the pixels with registration error as outliers for SR reconstruction. This is the general concept of robust SR reconstruction.

The mathematical expression to reject the certain pixel is to assign it zero weight. The other approach is to use a robust error function or M-estimator instead of the *L*2 norm.
The general expression of the cost function of the robust SR reconstruction is

$$I = \sum_{i=1}^{N} \sum_{j=1}^{M_i} w_{i,j} \rho(b_{i,j}^{T} . x - y_{i,j}) + \alpha\lambda(x) \ldots\ldots (6)$$

where, *wi,j*is a real value weight or a binary mask for the *j*-th pixel of the *i*-th observed image, and function *ρ*(*x*) is a robust error function. The key of robust super-resolution is how to design the weight and robust error function adaptively.
First, overview the pixel-weight approach. In the pixel-weight approach, zero weights or small weights are used to reject the outlier pixels, where the error function is the *L*2 norm.

Deblurring:

Since will not include the effect of sensor blur in the data model of the HR frame, and then the deblurring is necessary as a post processing step to improve the outputs HR further. Defining the estimated frame at time *t* as[5]

$$\hat{z}(t) = [\ldots, \hat{z}(x_j), \ldots]^{T} \ldots\ldots (7)$$

where *j* is the index of the spatial pixel array and *u*(*t*) as the unknown image of interest, we deblur the frame *z*(*t*) by a regularization approach:

$$\hat{u}(t) = \arg\min_{u} \| u(t) - G\hat{z}(t \quad \|_2^2 + \lambda C_R(\hat{u}(t)) \ldots\ldots (8)$$

where *G*is the blur matrix, *λ*(≤ 0) is the regularization parameter, and $C_R(u)$ is the regularization term. More specifically, we rely on our earlier work and employ the *bilateral total variation (BTV)* framework:

$$C_R(u(t)) = \sum_{v_1=-v}^{v} \sum_{v_2=-v}^{v} \eta^{|v_1|+|v_2|} \| u(t) - F_{x_1}^{v_1} F_{x_2}^{v_2} u(t) \|_1 \ldots\ldots (9)$$

where *η* is the smoothing parameter, *v* is the window size, and *Fv*1 *x*1 is the shift matrix that shifts *u*(*t*) *v*1-pixels along *x*1-axis.

In the present work, we use the above bilateral total Variation *(BTV)* regularization framework to deblur the upscaled sequences frame-by-frame, which is admittedly suboptimal[4].

In our work we have introduced a different regularization function called *Adaptive Kernel Total Variation (AKTV).* This framework can be extended to derive an algorithm that can simultaneously interpolate and deblur in one integrated step. This promising approach is part of our ongoing work and is outside the scope ofpaper.

Projection Onto Convex Set:

According to the method of POCS for super-resolution reconstruction, the space of estimated high-resolution solutions is restricted by a set of constraints (closed convex sets) which characterize desirable properties, such as fidelity to data, smoothness, sharpness etc., to be consistent in the final solution. For each set of convex constraints C_i, a projection operator T_i is defined. The problem is then reduced to iteratively locate, given a point in the high-resolution image space, the closest solution which intersects with all the given convex constraints, C_i. The convergence can be given as[6]:

$$X^{n+1} = T_i X^n$$

$$\Rightarrow X^{n+1} = T_m T_{m-1} \ldots T_2 T_1 X^n \quad \text{for } n = \square, 1, 2 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (10)$$

6. Motion Estimation Error

Motion estimation error in a super-resolution image reconstruction algorithm is an important part and it relates to the adjacent sub-pixel information between frames can be effectively utilized. The use of the image intensity of the inherent two-dimensional motion estimation (pore size, cover, reveal effects, etc.), the result of motion estimation error can be seen as noise. To effectively control not accurate motion estimation errors, need to judge these areas. This approach will be classified as POCS, can create noise on the error set of constraints as follows[6]:

$$C_M^k[n_1, n_2] = \{x[n_1, n_2]: |D(n_1, n_2)| \le \sigma_k\} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (11)$$

There for

$$D(n_1, n_2) = \sum_{p=-1}^{1} \sum_{q=-1}^{1} |X(a(n_1 + p) + MV_y, b(n_2 + q) + MV_x) - y(n_1 + p, n_2 + q)| \ldots \ldots \ldots \ldots (12)$$

σk is a reference frame fixed in the center of that neighborhood where the standard deviation, and its value reflects the size of the local characteristics of image; when σk is large the texture regions of image be detailed, and when σk is small the area will be smoothed. POCS established operator as

$$P_{d(n_2, n_2)} = \begin{matrix} \text{internal repair}, |D(n_1 n_2)| < \sigma_k(n_1, n_2) \\ \text{directional interpolation}, |D(n_1, n_2)| > \sigma_k(n_1, n_2) \end{matrix} \ldots (13)$$

When the motion estimation of the projection error D (n1, n2)>σk (n1, n2), indicating the motion estimation error is larger at this time, while maintaining image quality as much as possible, the region-based direction of interpolation will repair.

Anaglyphic Process:

A frame-compatible stereogram where the left- and right-eye images are color-coded and combined into a single image.
When the combined image is viewed through glasses with lenses of the corresponding colors, a three-dimensional image is perceived. Different complementary color combinations have been used over the years, but the most common are red and cyan, a shade of blue-green. Anaglyphs may be projected, displayed on a monitor, or produced in print[8].
As can seen, the concept behind a stereo picture is to present slightly offset views of the scene to each eye to trick it into seeing depth. In the real world, we do this with two eyes viewing the scene from slightly different angles. To make a stereo picture, the challenge is not only to display the two different views on the same screen, but also to show just one of those views to each eye. It turns out that this is hard to do. One of the earliest workable methods was the anaglyph process, using the familiar red and cyan glasses.
Figure (2) shows a typical anaglyph picture. If you have a pair of red and cyan anaglyph glasses lying around, you can try them out on this picture. The anaglyph process merges the two offset views into a single image, with one view colored cyan and the other view color red. The red and cyan lenses of the anaglyph glasses then allow each eye to see its intended view while blocking the other eye.

The problems with anaglyph are ghosting, color degradation, and retinal rivalry (different brightness to each eye). It's just not a very good way to present separate views to each eye. In movie theatres, high-tech polarizing display systems and glasses are used for their much better quality and low cost for a mass audience. However, in a production studio active shutter systems are most common. An active shutter system has rapidly blinking LCD glasses synchronized to a display that is rapidly flicking between the left and right eye views[9].
Although anaglyphs provide an inexpensive means of 3D visualization, there are some drawbacks related to their use. The primary disadvantage is that, unlike the shuttered or polarizing glasses techniques, the use of a fixed pair of color filters does not allow the generation of arbitrary ghost-free full-color 3D visualizations. Each pixel in an anaglyph gets its color from both left and right images and the result is grey for equal red/cyan combination and yellow for the equal red/green combination. It is a challenging task to accurately display the colors of the original stereo image into a single pixel in the anaglyph image. The problems which may cause degradation of the quality of the anaglyph are, ghosting or crosstalk, retinal rivalry or called binocular rivalry, and color merging. Ghosting reduces, and might sometimes prevent the viewer perceiving depth [10].
This phenomenon occurs when the eye sees the image it is supposed to see and also a part of the other image which should be totally blocked from the sight of this eye. This may result in a double image in some parts of the scene. Retinal rivalry or binocular rivalry occurs when all colors are not displayed by the display device. Color merging is the phenomenon which occurs when color components of the left and right images are transferred to the same color .
The pair images should be offset into a proper position before the combination. There are two types of pair images: calibrated and un-calibrated images, where the separation angle between two cameras should be taken into account [11].

Figure (2), Anaglyph picture

Stereoscopic Image Processing for 3D Monitoring:

Once again, let's refer back to our color analogy. There are three levels of color monitoring: basic control is watching lowlights and highlights, making sure images are not burned out; intermediate color control just requires an all purpose monitor; and advanced color control requires a color-calibrated monitor, light-controlled environment, or the trained eyes of an expert colorist or DP. Otherwise, reference patterns and a vectorscope would do the job, even with a colorblind technician. Beyond this, onset LUT management will allow the crew to make sure the tweaked image will match the desired artistic effect[8].

Experimental Result and Analysis :

First, translation motion for each low resolution frames and set resolution factor, and implemented them by function to all frames and space invariant.

Estimates (Robustly) the blurred high resolution as a median of the LR images after up sampling and shifting to correct position.

An efficient implementation is performed here in which we note that the high resolution image consists of the median of all the LR images which have the same displacement value. That is, the LR images are partitioned to unique groups, each group have equal displacement values in the HR image. The median of each group is then up sampled and shifted into the HR image.

In some cases, not all displacements exist. This leaves us with undetermined cases. In this case some other interpolation method needs to be implemented. In this implementation we "fill" the holes using a spatial median filter.

Computes the gradient back-projection for the fast deblurring and interpolation method. This function implements the gradient of the level one norm between the blurred version of the current deblurred HR estimate and the original blurred HR estimate created by the median and shift method.

HR reconstruction of the *i-th* frame given the whole image sequence. Eventually, we evaluate the conditional expectation of $SR(v_i)(x)$ for any $x \in \Psi$, given the information of all of the frames $\{u_j\}$ for $1<j<k$:[6]

$$E[SR(v_i)(x)\{u_j\}_{1\le j\le k}] = \frac{1}{G(i)} \sum_{i\le j\le k} g(|i-j|) \times E[SR(v_i)(x)|u_j] \quad \ldots (14)$$

$$G(i) = \sum_{i=j}^{k} g(|i-j|) \quad \ldots (15)$$

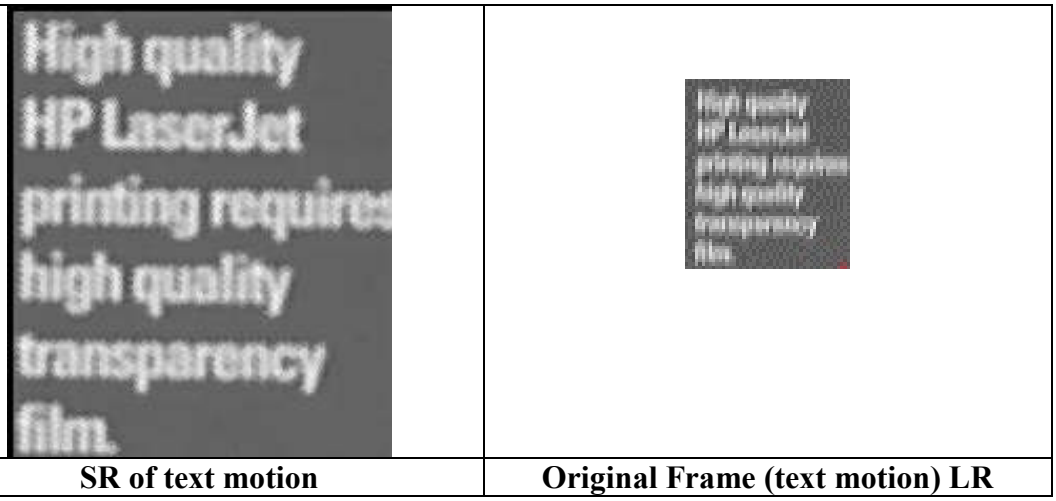
where *g* is a decaying function of |*i-j*|, and *G* is a normalization factor. The expression *g(|i-j|)* in this equation represents the temporal confidence on the expectations computed for each of the various frames, *j*, which has been taken into account in reconstructing the HR image $SR(v_i)$. In the experiments reported below, we have assumed that each of the frames in hand are equally likely useful in producing the HR details of the *i-th* frame. Hence, we have taken g to be a box-function of large enough support which yields g(|i-j|)= 1. As a result,

$$E[SR(v_i)(x)\{u_j\}_{1\le j\le k}] = \frac{1}{k}\sum_{j=i}^{k} E[SR(v_i)(x)|u_j] \dots\dots\dots\dots\dots (16)$$

By taken a frame from motion sequence of video, we've set resolution factor and implemented with HPSF function, as in equation blow

$$\forall_\omega 0 < \left|1 - \frac{HPSF(\omega)HAUX(\omega)}{G}\right| < 1 \dots\dots\dots\dots\dots\dots\dots\dots (17)$$

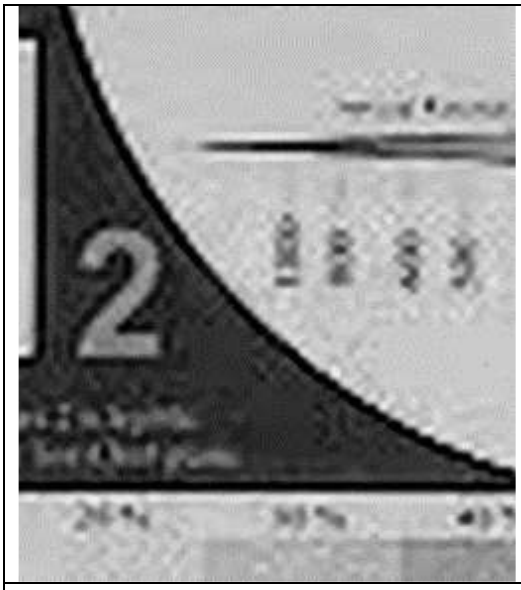
And deblured adds in eq.(8, 9) the LR frames, we got the results as below which is a frame of sequences of image as shown in figure (3).



SR of text motion	Original Frame (text motion) LR

Figure (3), text motion in LR and SR

Other extracted frame of motion is "adds motion" as shown in figure (4), with less vector than the text motion.



SR of adds motion	Original Frame (adds motion" LR

Figure (4), adds motion frame in LR and SR

For more results and experience in the proposed work, we've take consideration of different motion images, and one of them is "car motion" as shown in figure (5).

Car motion in SR	Car motion in LR

Figure (5), car motion in LR and SR

Now, as shown in below, will convert the last results of processed frames (from Low-resolution to Super-resolution) to an anaglyph method for proposing of 3D-technology.

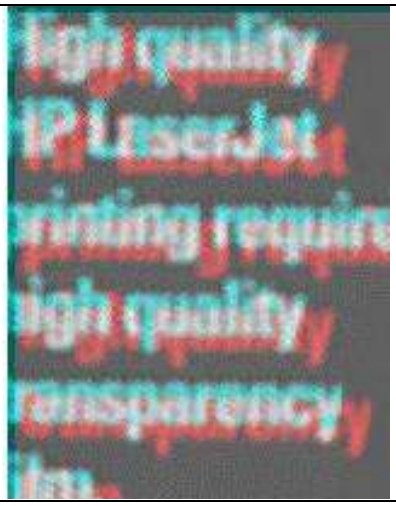
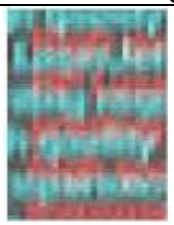
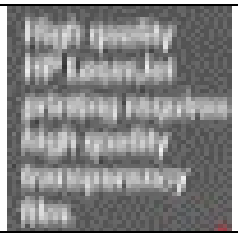
	
Text motion with an Anaglyph in SR	SR of text motion
	
Text motion with anaglyph in LR	LR of text motion

Figure (6) LR and SR of "text motion" frame with an anaglyph 3D method

As shown above the proposed anaglyph 3D method in super high resolution has better results than one has implemented in low resolution. Another test has take in consideration for different image as in the "adds motion" as shown in figure (7).

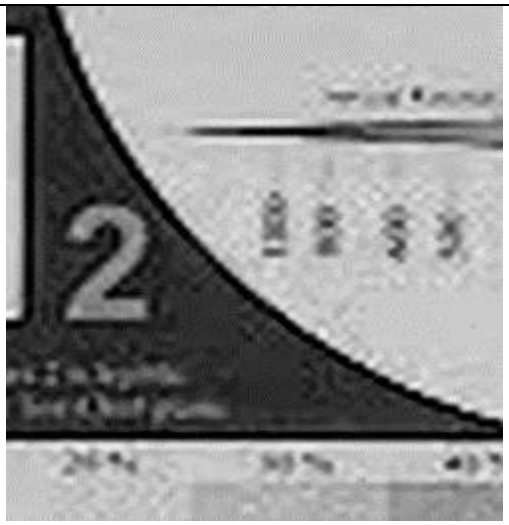
	
Adds motion with an Anaglyph SR	SR of adds motion
	
adds motion with anaglyph in LR	LR of adds motion

Figure (7), LR and SR of "adds motion" frame with an anaglyph 3D method

Toy motion in SR Toy motion in Anaglyph SR

Figure (8) SR of "toy motion" frame with an anaglyph 3D method

Car motion in SR Car motion in Anaglyph SR

Figure (9) SR of "car motion" frame with an anaglyph 3D method

In Table (1) below, different results efficiently calculated between each image in both cases in low resolution and super resolution, and before and after an anaglyph applies.

Table (1)

PSNR Peak Signal to Noise Ratio	MSE Mean Square Error	Image
22.8416	338	Text Motion (Low Super Resolution)
23.5369	288	Text Motion (High Super Resolution)
23.6884	278.1250	Adds motion (Low Super Resolution)
21.4493	465.7500	Adds motion (High Super Resolution)
22.8416	338	Car motion (Low Super Resolution)
24.0975	253.1250	Car motion (High Super Resolution)
22.8416	338	Toy motion (Low Super Resolution)
23.9065	264.5000	Toy motion (High Super Resolution)

MSE and PSNR calculations with different images in both SR and LR

As seen in table, each frame extracted from video has calculated both of MSE and PSNR compared original image with image has applied an anaglyph technology. The MSE results shown that error on low super resolution has same because it has didn't affected which this came by not applying any kind of method, in high super resolution the images has affected, for that the MSE values has changed and differenced from one image to other based on size and resolution of image.
The PSNR values for low super resolution images has compared with high resolution image, and as mentioned in MSE results, the results has been differenced from one image to other based on size and resolution of image and what the super resolution method came with effects on image.

Conclusion :

Although there are a range of other stereoscopic display technologies available that produce much better 3D image quality than the anaglyph 3D method (e.g. polarized, shutter glasses, and Infitec), the anaglyph 3d method remains widely used because of its simplicity, low cost, and compatibility with all full-color displays and prints.
This work has focus on anaglyphic 3-D images can be product mass movie enhancement in 3D based on super resolution technology. The work has also revealed that there is considerable variation in the amount of anaglyphic exhibited by different display, different movie, and different anaglyph colors in case if using different glasses color. Compared to works that has only considered red/cyan anaglyph glasses, this paper has extended the work to include red/yellow which are now also in common usage.

References:

1- Aggelos K. Katsaggelos, Rafael Molina and Javier Mateos; *"Super Resolution of Image and Video"*; Morgan & Claypool Publishers; 2007.
2- Bernard Mendiburu, YuvsPupulin, Steve Schklair; *"3D TV and 3D Cinema, Tools and Processes for Creative Stereoscopy"*; Focal Press;2011.
3- Styles, P.J. The Production Of Anaglyphs From Spot-HRV Panchromatic Data For Geomorphological Mapping. 1990.
4- BirBhanu, Ju Han; *"Human Recognition at a distance in video"*; Springer Press; 2011.
5- Lisa Yount; *"Virtual Reality"*; "The Luncent Library of Science and Technology"; 2005.
6- PeymanMilanfar; *"Super Resolution Imaging"*; CRC Press; 2010.
7- Steve Wright; *"Digital Compositing for Film and Video, Featuring new content on 3D compositing, Stereo Compositing, and Multi-pass CGI Compositing"*; Focal Press; 2010.
8- SubhasisChaudhuri, Manjunath V. Joshi; *"Motion Free Super Resolution"*; Springer Press; 2005.
9- SubhasisChaudhuri; *"Super Resolution Imaging"*Springer Press; 2002.
10- Yei-Yu, Y. and D.S. Louis,*"Limits of fusion and depth judgment in stereoscopic color displays"*;Hum. Factors, 1990. 32(1): p. 45-60.
11- Sanders, W. and D.F. McAllister. *"ProducingAnaglyphs from Synthetic Images"*;2003.