6. Intermediate viewpoint generation

This chapter concerns the generation of intermediate viewpoint images. The reasons why such images might be required are briefly discussed in Section 6.1.

Intermediate viewpoint images generated using left and right images and disparity data correspond to camera positions on a trajectory between the positions left and right camera respectively. The nature of this trajectory between left and right camera is slightly different for parallel or converging configurations. This is discussed in Section 6.2.

The basics of the generation of an intermediate viewpoint image are dealt with in Section 6.3. Because occlusion implies that some points will only be visible in either left or right source image appropriate handling of occlusion will be required. This is discussed in Section 6.4.

In Section 6.5 intermediate viewpoint generation is looked at from a system standpoint.

Finally in section 6.6 experimental results are presented using disparity fields generated using a variety of the disparity estimation techniques presented in earlier chapters..

6.1 Application of intermediate viewpoints

When viewers first experience the added realism of stereoscopic video one of the first reactions is invariably to "walk round" the scene, to look at the scene from another point of view as illustrated in Figure 6.1 where two alternative viewer positions are sketched along with the central viewer position which corresponds to the camera position.

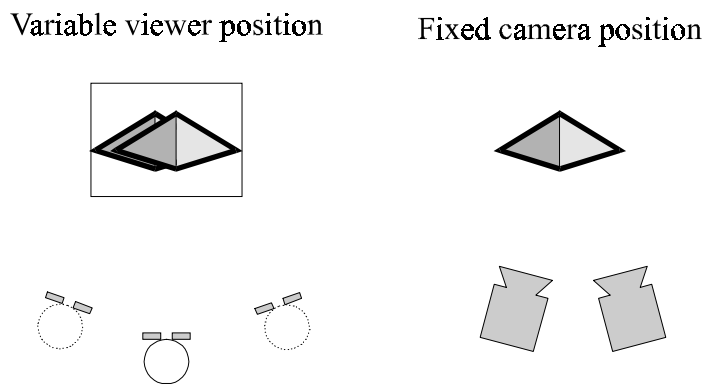

Figure 6.1: Variable viewer position.

With stereoscopic material this looking round does not work. The displayed stereoscopic material is dictated by the position of the stereoscopic camera which recorded it. The position of the viewer has no influence whatsoever on what is seen.

What happens when the viewer moves is that the scene appears to move in counter sense. This is not a natural experience and thus it, in a sense, counteracts the realism of stereoscopic video.

A degree of added realism could be added and/or restored if it were possible to synthesise images, to be presented to the left and right eye of the viewer, which are appropriate to the position of the viewer relative to the display screen. Given a stereoscopic video sequence and the corresponding disparity vector fields it is in fact possible to generate intermediate views (as described in the following Sections). If the position of the viewer relative to the screen is then mapped to a left/right pair of intermediate camera positions for which views are synthesised and presented to the viewer in real time then the realism of stereoscopic video can be considerably enhanced.

The importance of intermediate viewpoints will vary from application to application.

In a stereoscopic cinema the viewers would be sitting in a static position, their assigned seat. Given the distance to the screen the viewer would hardly expect a perspective change to accompany head motion. It might even be undesirable if viewers at the left and right of the cinema see different perspective views, probably they would want to the composition as made by the director of the film. Generating appropriate intermediate views for each of several hundred cinema visitors could be technically problematic.

An application where intermediate viewpoints could be important is a stereoscopic tele-conferencing / tele-presence setup where the viewer is sitting close to a relatively small screen. When the viewer moves his head he will expect a perspective change. If the objective of the stereoscopic communications link is to co-operatively discuss and virtually handle some object then the ability to view is from varying positions may well be very desirable and form a large part of the added value of a stereoscopic link as opposed to a monoscopic link.

6.2 Intermediate viewpoint camera trajectory

The nature of the intermediate camera positions which can be synthesised using left/right data and disparity fields varies slightly depending on whether the source material was recorded using a parallel or a converging camera configuration.

For the parallel camera configuration the iso-disparity contours run parallel to the X axis. Examination of equation (2.6) shows that the magnitude of the disparity and the baseline distance which separates the cameras are directly proportional. Intermediate views generated using disparity data and left/right image source material correspond to camera positions on the X axis along the baseline, ranging

from the left camera position (-*B*/2,0,0) to the right camera position (*B*/2,0,0) as illustrated in Figure 6.2.

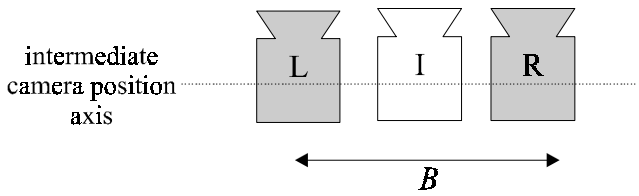

Figure 6.2: Intermediate viewpoint for parallel cameras.

If a number of intermediate viewpoints are calculated for a stereoscopic still image pair and these viewpoints are displayed monoscopically in a time-sequential fashion then this is perceived as the scene sliding from left to right.

For parallel camera positions the situation, as illustrated in Figure 6.3 is a little different.

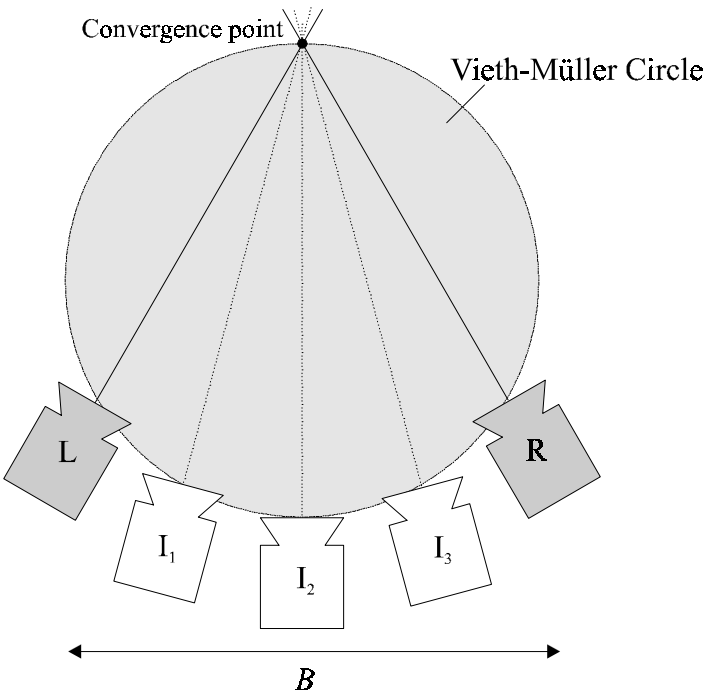

Figure 6.3: Intermediate viewpoint for converging cameras.

Obviously all points which have zero disparity will be located in the same position in all viewpoints. This means that the intermediate camera configuration must share the same Vieth-Müller circle as the original camera configuration. The focal points of the camera lenses must lie along that same Vieth-Müller circle and the camera convergence point must still correspond to the furthest point along that circle.

So the intermediate trajectory for converging cameras corresponds to an arc along the Vieth-Müller circle, starting at the original left camera position and ending at the right camera position.

If a number of intermediate viewpoints are now calculated for a stereoscopic still image pair and these viewpoints are again displayed monoscopically in a time-

sequential fashion then this is perceived as a rotation of the scene. The accompanying motion parallax can give a strong impression of depth even in monoscopic display.

6.3 Intermediate viewpoint computation

Intermediate viewpoint generation for stereoscopic images is the generation of an image which corresponds to the generation of an image as it would have been perceived by an imaginary camera in a position somewhere between the left and right cameras. The correspondence between points in the left and right images is given by disparity which can be modelled as a horizontal shift (Chapter 2).

Mathematically intermediate viewpoint generation comes down to calculating the image which would have been perceived by an imaginary intermediate viewpoint camera placed somewhere along a parametrised trajectory between left and right cameras.

The left image is given by:

$$S_L(x,y) \quad 0 \le x < W_{pix}, \quad 0 \le y < H_{pix}$$

and similarly the right image by:

$$S_R(x,y) \quad 0 \le x < W_{pix}, \quad 0 \le y < H_{pix}$$

For those points in the left image which are not occluded the left and right disparities are:

$$d_l(x,y) \quad 0 \le x < W_{pix}, \quad 0 \le y < H_{pix},$$

$$d_r(x,y) \quad 0 \le x < W_{pix}, \quad 0 \le y < H_{pix}$$

The parameter which maps the trajectory is s. $s = 0$ corresponds to the left camera viewpoint while $s = 1$ corresponds to the right camera viewpoint. For points which are not occluded in the left image the intermediate viewpoint values can be extrapolated from the left image by:

$$S_{I,L}(x + d_l(x,y) \cdot s, y) = S_L(x,y).$$

Similarly for point which are not occluded in the right image intermediate viewpoint values can be extrapolated from the right image by:

$$S_{I,R}(x - d_r(x,y) \cdot (1-s), y) = S_R(x,y).$$

As such these extrapolated values can be used directly as intermediate viewpoint values. Choosing between the two based on which source camera is nearer:

$$S_{I,E}(x,y) = \begin{cases} S_{I,L} & \text{if } 0 \le \mathrm{s} < 0.5 \\ S_{I,R} & \text{if } 0.5 \le \mathrm{s} < 1 \end{cases}.$$

Weighted averaging of the two extrapolated images is also a possibility:

$$S_{I,A}(x,y) = (1-s)\cdot S_L(x,y) + s\cdot S_R(x,y).$$

Note that occlusion is not yet accounted for in this description. The handling of occlusion is discussed in Section 6.4.

Which technique to use, selection or averaging is a matter of viewer preference and of the quality of the source material. If the source material has been recorded by badly balanced cameras then the luminance difference between S_L and S_R may be large. If the viewpoint gradually changes and crosses the s = 0.5 boundary then the luminance of the presented intermediate viewpoint image will also suddenly change. This is not a desirable effect. Using weighted averaging makes the change gradual and considerably less noticeable.

On the other hand the meaning of weighted averaging of two extrapolations is debatable. An intermediate image (luminance) value is calculated which may be present in neither left or right image. If the quality of left and right vector fields is not very good then nonsense values may be generated.

On balance the weighted averaging is probably preferable if the cameras are not perfectly balanced and disparity fields of sufficient quality are available.

6.4 Handling occlusion

In scenes where objects are interposed some degree of occlusion will occur. Small parts of the scene which are visible to the left camera are not visible to the right camera and vice versa. A small example is illustrated in Figure 6.4.

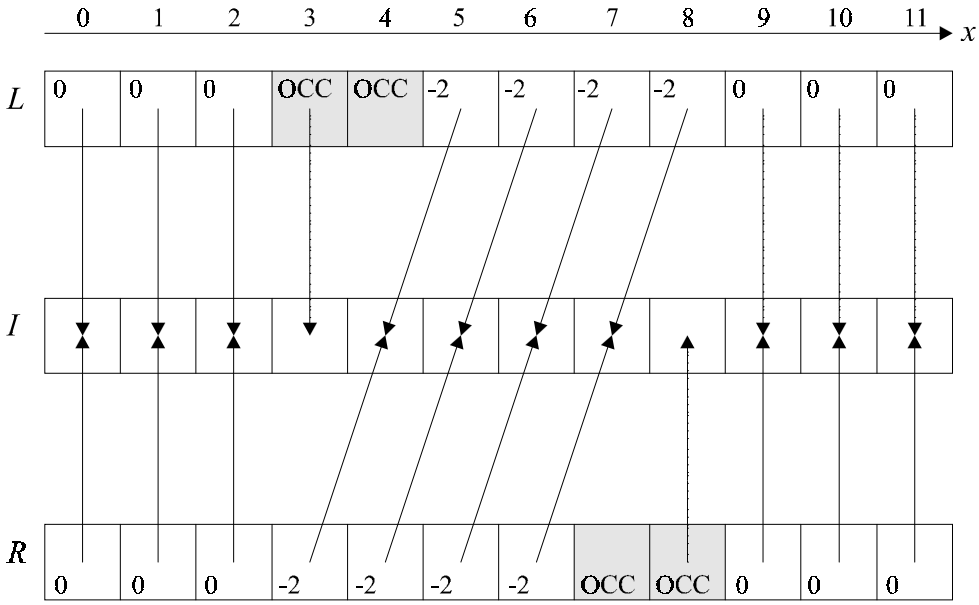

Figure 6.4: Occlusion in intermediate viewpoint generation.

The (single scan-line from a) scene illustrated has a background at zero disparity. There is one object in the foreground at x = 5 to x = 8 in the left scan-line and x = 3

to $x = 6$ in the right-scan line. This object has a disparity of -2. The intermediate viewpoint line is generated for the middle viewpoint position, i.e. $s = 0.5$. Therefore the disparity vectors which point from the left/right source lines to the intermediate line are halved in magnitude. The points at $x \in \{0,1,2,4,5,6,7,9,10,11\}$ can be either interpolated or extrapolated. However at $x = 3$ and $x = 8$ a problem occurs. These two points correspond to occlusions at the left and right sides of the occluding object: left and right occlusions.

At $x = 3$ a left occlusion occurs. Two points on the left scan-line have no correspondence along the right scan-line. This is an occlusion of width two corresponding to a disparity difference of two. On the intermediate scan line where the disparity is halved the size of the occlusion is also reduced to one position which is not pointed to by a (scaled) disparity vector. Such a locations which is not pointed to is referred to as a hole. The value to fill in for this hole should obviously be extrapolated from the left scan-line. Obviously it is desired to fill in a piece of the background and thus the extrapolation should be performed using the left disparity vector used for the location to the left of the hole (in the figure this vector has zero magnitude).

Similarly at $x = 8$ on the intermediate scan-line there is a hole caused by a right occlusion. Here the background data should be extrapolated from the right scan-line. The disparity vector corresponding to the background disparity can now be found by using the right disparity vector used to the right of the hole.

So when an intermediate viewpoint scan-line is generated two kinds of holes will occur: holes caused by left occlusions and holes caused by right occlusions.

- Holes caused by left occlusions have a higher (interpolated) disparity to the left of the hole than to the right of the hole. The disparity vector directly to the left of the hole should be used to extrapolate values to fill in the hole from the left scan-line.
- Holes caused by right occlusions have a lower (interpolated) disparity to the left of the hole than to the right of the hole. The disparity vector directly to the right of the hole should be used to extrapolate values to fill in the hole from the right scan-line.

If left and right vector fields are not consistent then there is the possibility of multiple assignments: more than one interpolated disparity vector from each scan-line pointing to a single position along the intermediate scan-line. To alleviate this it is possible to discard inconsistent disparity vectors, i.e. disparity vectors for which

$$d_r(x + d_l(x,y), y) = d_l(x,y)$$

does not hold. If there are many such vectors along a scan-line and they are discarded then it is likely that a third class of hole will be generated.

- Holes which have an equal (interpolated) disparity to the left of the hole and to the right of the hole can be filled in either by interpolation or extrapolation from both left and right scan-lines.

In principle only one correct disparity field is required is required for intermediate viewpoint generation. Supposing for instance that the left disparity field is available and areas of occlusion have been correctly flagged in this field. The vectors from this disparity field link all non-occluded pixels in left and right images. Extrapolation or interpolation is possible for these pixels. The holes which are then left can be taken care of using the rules described above.

6.5 System view of intermediate viewpoint generation

Given left and right input images and a viewpoint parameter, *s*, two actions need to be performed: disparity estimation and the actual viewpoint generation itself. In practice intermediate viewpoint generation for stereoscopic video will be performed in the context of some kind of communications system. Stereoscopic video comes from a signal source, is coded and transmitted across a channel, is decoded at receiver side and is viewed by one or more viewers.

Disparity estimation can now be performed at either source side, before encoding - transmission - decoding, or at receiver side. Equally the intermediate viewpoint generation itself can be performed at sender or receiver side.

The scenario in which disparity estimation and viewpoint generation are both performed at sender side is illustrated in Figure 6.5.

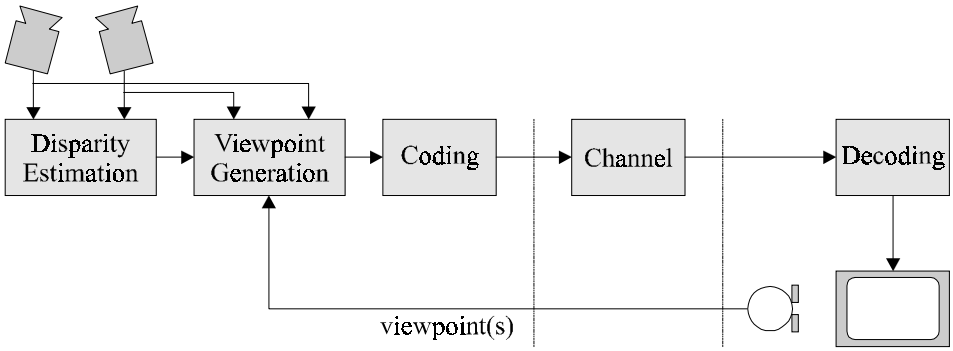

Figure 6.5: Estimation and generation at sender side.

In this scenario all processing involved in intermediate viewpoint generation is performed at sender side. Disparity estimation performed at sender side can be performed on source material which has not been affected by coding artefacts. Performing viewpoint generation at sender side is not so advantageous. Two complete viewpoint sequences have to be transmitted for each viewer. This makes this configuration very bit rate intensive for anything but one-on-one communication. Also the delay between the viewer taking up a certain position and this position being reflected in the images displayed is large because it is the sum of

several component delays: transmission of position over the channel - viewpoint generation - coding of viewpoint images - transmission over channel - decoding of viewpoint images. This cumulative delay seriously hampers natural perception. If a viewer moves his head and the viewpoint changes half a second later then this is not acceptable [**ask Sigmund Pastoor for a proper reference**].

Because of the possibility of multiple viewers at multiple positions and the delay inherent in viewpoint generation at sender side it is better to separate disparity estimation and intermediate viewpoint generation as illustrated in Figure 6.6.

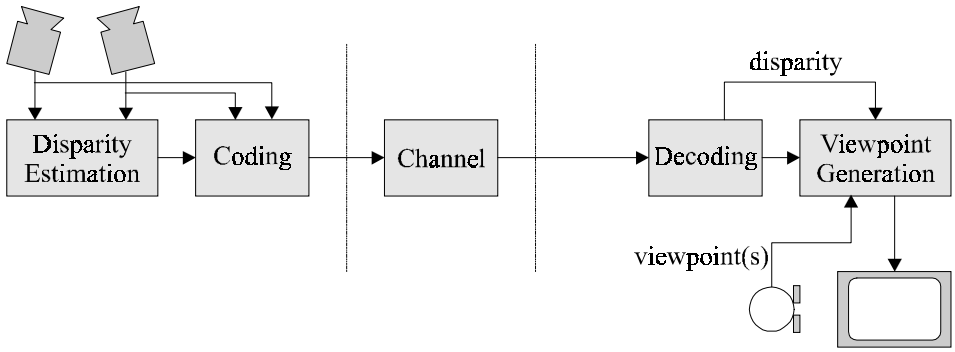

Figure 6.6: Estimation sender side / generation receiver side.

Disparity estimation is still performed at sender side on source material uncorrupted by coding artefacts. The resulting dense disparity field is transmitted across the channel. The intermediate viewpoint generation is performed at receiver side using the decoded images and disparity field. The only delay which now occurs between viewer positioning and corresponding viewpoint presentation is the delay incurred by performing the viewpoint generation itself.

The desirability of the second configuration over the first suggests the separation of disparity estimation and intermediate viewpoint generation. In a physical as well as a systematic sense. The disparity vector field becomes a given to the intermediate viewpoint generation unit which uses these disparity vectors, assuming them to be correct. The quality of the images perceived by the viewer at receiver side is now directly dependent on the quality of the disparity field estimated (an coded) at sender side.

The situation so generated is similar in nature to the one present with the S-MPEG codec (and MPEG codecs in general) where, within the constraints of bit rate and syntax, there is considerable variation in quality possible by careful design of some of the decisions and strategies in the encoder. In fact the influence of the quality of the estimated disparity field on the generated viewpoint images is probably even higher because inadequacies in the estimated disparity field show up directly as artefacts. There is no additional error signal (like in coding) to cover up the errors, no matter how high the bit rate.

Intermediate viewpoint generation is a severe test of any disparity estimator.

6.6 The necessity of accurate disparity fields

As intermediate viewpoint generation occurs after disparity estimation it is important that accurate disparity vectors be supplied to the intermediate viewpoint generator. Spurious match vectors which do not correspond to true disparity will show up the problem shown in Figure 6.7.

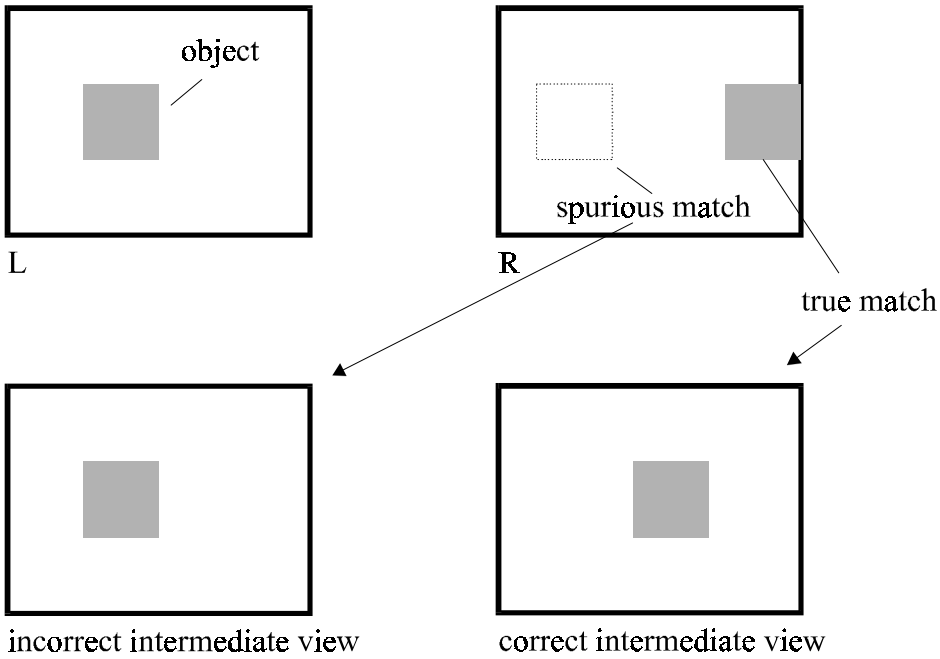

Figure 6.7: Necessity of accurate match.

A disparity vector based on a spurious match may well relate similar (blocks of) pixels in the left and right view. When such a vector is used for intermediate viewpoint generation material will be generated in an incorrect position, thus generating clear artefacts.

6.7 Experiments

The disparity estimators discussed in the previous chapters were used to estimate disparity between the left and right view images of the last frame of the three variants of the mirror sequence. The corresponding intermediate views for these are present in the mirror sequence itself.

In Figure 6.8, Figure 6.9 and Figure 6.10 the difference between the generated intermediate views and the actual intermediate views is plotted in SNR terms against the viewpoint *s*.

The relative ranking of the results is as with the accuracy figures of previous chapters. The DSI based estimator has the additional advantage that it is capable of detecting occlusion and passing this information on to the intermediate viewpoint generation routine. This is reflected in the fact that the curve for the DSI based

estimator does not show as strong a dip towards the middle (s = 0.5) viewpoint position as do the other estimators.

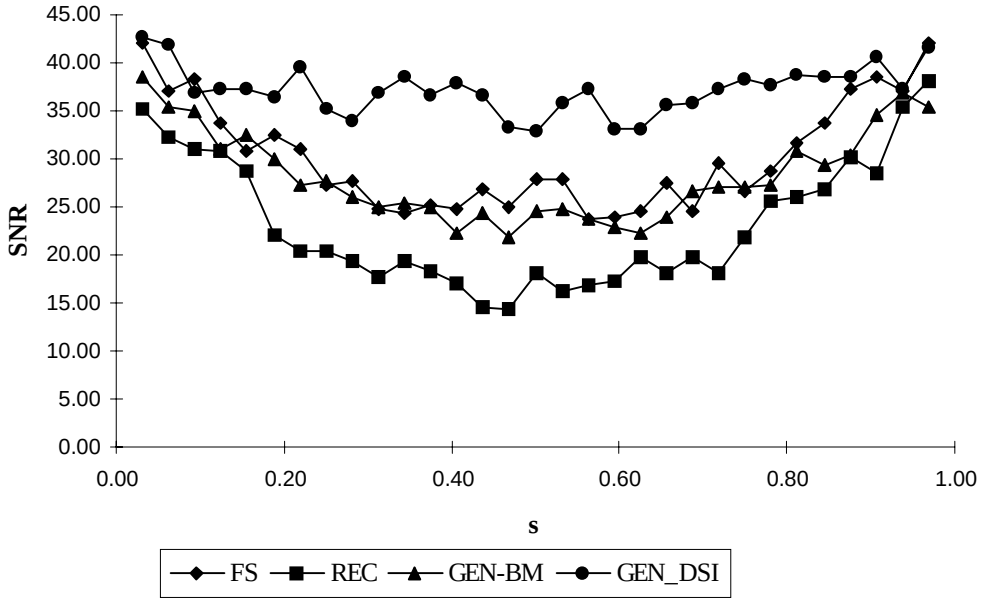

Figure 6.8: Intermediate viewpoint generation on miror_I.

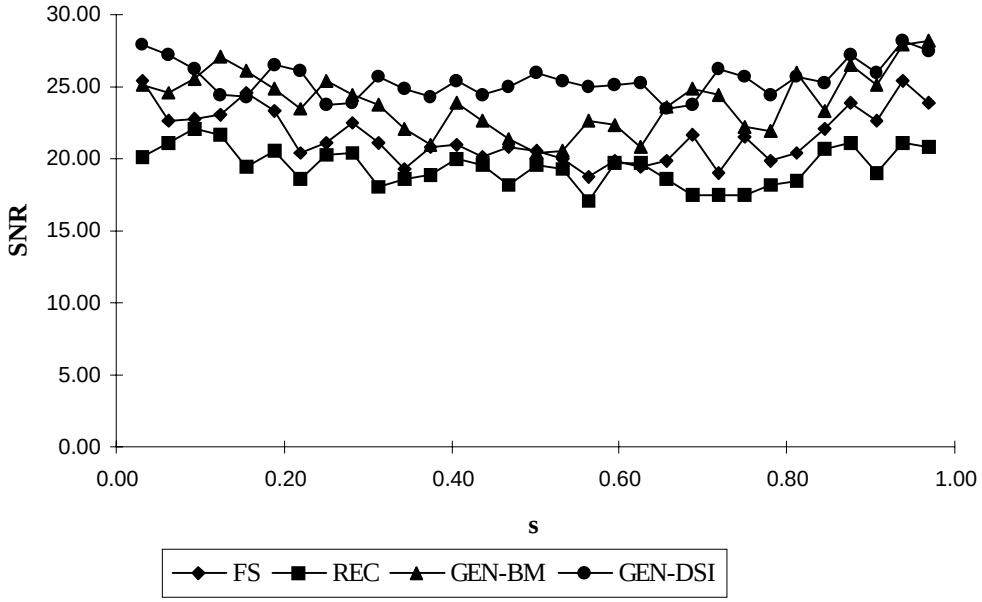

Figure 6.9: Intermediate viewpoint generation on miror_30.

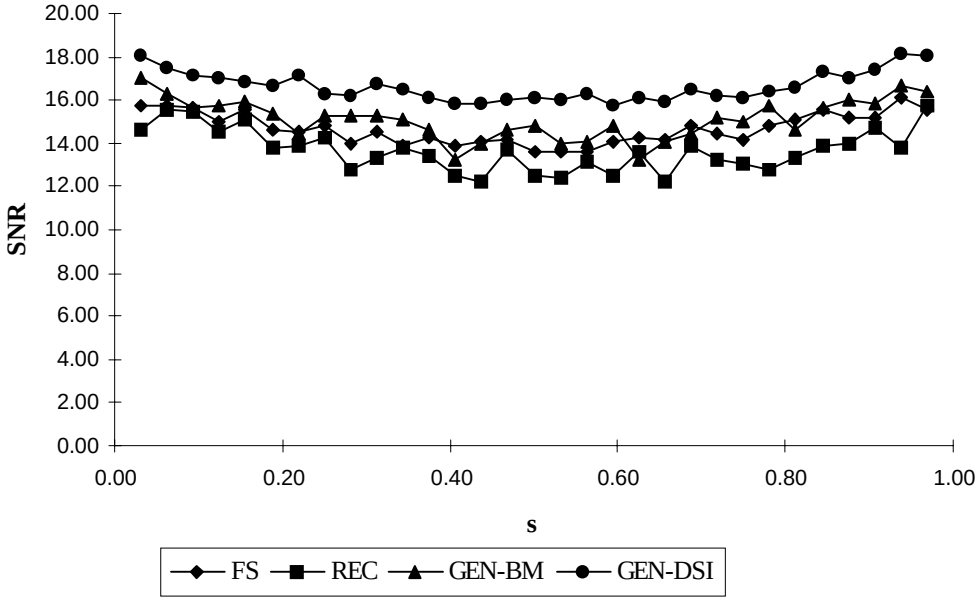

Figure 6.10: Intermediate viewpoint generation on miror_20.