An efficient disparity map format for real time interpolation in multi viewpoint stereoscopic video systems

**Keywords:** disparity map format, multi viewpoint system, stereo image interpolation

*André Redert, Emile Hendriks*

Information Theory Group, Department of Electrical Engineering
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
email: {andre,emile}@it.et.tudelft.nl
phone: +31 15 278 6269, fax: +31 15 278 1843

**ABSTRACT**

In this paper we present an efficient disparity map format that has been developed in the European ACTS project PANORAMA for stereoscopic multi viewpoint video systems. This new map, the chain map, has several advantages compared to the usual disparity vector fields. It is a very efficient, single bidirectional map, with explicit indication of occlusions and normal/contracted objects, there are no upper or lower bounds on disparity values and a disparity offset is not needed.

For real time video applications, the chain map allows for low complexity hardware implementations of system parts. We present disparity estimation, image interpolation and filtering algorithms with low hardware complexity. With teleconferencing images high quality results are obtained.

**1. INTRODUCTION**

Stereoscopic multi viewpoint systems have been considered widely for 3D video communication systems [1-6]. These systems provide the sensation of depth and motion parallax. Figure 1 shows a multi viewpoint system currently being implemented in the European PANORAMA project [1]. First stereo images are captured and analysed in a disparity estimation step [2,3], resulting in a disparity map. The map indicates the correspondences between left and right images. Finally a synthesis step, image interpolation [4,6], is applied based on the measured position of the viewer.

The format of the disparity map has a strong influence on the system complexity. In video communication applications the feasibility of hardware implementation is important, so a simple and efficient disparity map format is required.

Recently, we developed a new and efficient disparity map format [5] for the PANORAMA multi viewpoint system, as an alternative to the commonly used vector fields [2,3,6].

In this paper we will show that this map is not only efficient but also lowers the complexity of hardware implementation.

First we describe which disparity maps are considered, followed by an introduction in the chain map. Then we present new disparity estimation and image interpolation algorithms that are especially suited for the chain map and have low hardware complexity. We show that filter operations can be applied easily in the chain map domain. Finally we give experimental results and conclusions.

**2. DISPARITY MAPS AND CONSTRAINTS**

A disparity map indicates which pixels in left and right images correspond to the same point of the recorded scene. We consider disparity maps under four constraints:

- **The map has pixel resolution and accuracy.**
- **Disparity has a horizontal component only.** All pixels in one scanline of the left image should correspond only to right image pixels on the same scanline.
- **Ordering constraint.** The order of objects from left-to-right should be the same in the left and right images.
- **In occluded areas, the disparity map contains no pseudo disparity information.** At occlusion areas, disparity vectors do not exist. Still one can generate pseudo disparity vectors, pointing from the visible part in one image to the imaginary counterpart in the other image. We exclude their presence in the disparity map.

Figure 2 shows a single scanline of the stereo image pair, together with a ‘scanline’ of the disparity map in two representations. On the left, pixel pairs are given by arrows in the traject representation. The arrows indicate a traject that the individual pixels follow when the camera position moves from left to right in a continuous way. On the right, pixel pairs are indicated...
by single points lying on a path in a two dimensional matching space [2,6].

Figure 2: A scanline of the stereo image pair and the disparity map
Scene object surfaces that are normal to the camera optical axes result in projections of the same size in left and right image. We call this normal objects. When an object surface is not normal to the camera axes, the projections in left and right images have different size. We call this contractions. We define a left contraction as an object that is larger in the left image than in the right image. An object that is visible in the left image, but not in the right image, is defined as a left occlusion.

3. A NEW DISPARITY MAP FORMAT: THE CHAIN MAP

The usual format for disparity maps is the vector field [2,3,6]. Figure 3 shows a scanline of the right-to-left and left-to-right vector field formats of the disparity map in Figure 2.

Figure 3: A scanline of the disparity vector field format
The advantage of a vector field is the complete freedom of disparity values, e.g. no ordering constraint.

But the vector field has several disadvantages. First, the finite number of bits assigned to each disparity value restricts disparities to a range. Secondly, an offset or scaling may be necessary to make full use of the available range. Finally, a single left-to-right vector field is unable to indicate the difference between a right occlusion and a right contraction. In both cases the disparity vectors are diverging, leaving pixels in the right image unpaired. This problem can be solved by using explicit occlusion fields or both vector fields simultaneously [6], but that introduces much redundancy.

We measure the number of bits in the disparity map format relative to the number of pixels in a single image from the stereo pair. Usually a disparity vector field contains 8 bit per pixel. Using both fields would result in a total format of 16 bit per pixel.

In [5] we introduced a new disparity map format, the chain map. The map is designed for efficient representation of all disparity maps under the constraints of section 2. Here we give a brief description of the chain map.

Figure 4 shows the chain map format of one scanline of a disparity map. For clearness, the traject and path representations are included in the figure.

Figure 4: A scanline of the chain map format
Starting at the first pixel in the left and right image scanlines (bottom-left corner of the disparity path), we indicate whether a disparity vector is present between the two pixels (M, a matching pair) or not (M). Then we follow the disparity path until the edge in both scanlines is reached. Each 1-pixel step in one of the two scanlines corresponds to an L or an R in the chain map. At each visited node in the path the presence of a match corresponds to an M in the chain map.

If N is the number of pixels on one scanline, then every scanline of the chain map consists of 2N-1 match bits and 2N-2 step bits. The total number of bits required for one scanline is thus 4N-3, which is slightly below 4 bits per pixel.

Advantages of the chain map are

- One bidirectional disparity map
- Explicit indication of occluded, contracted and normal objects
- No upper/lower bounds or offsets on disparities
- Only 4 bits per pixel

The chain map match bits correspond to a segmentation of the left and right images into objects and occlusions. Therefore they are hard to determine accurately.

If the match bits are discarded a reduced chain map is obtained of only 2 bit per pixel. This reduction lowers the complexity of the disparity estimator, since it does not have to generate the match bits anymore. The reduced chain map requires a special interpolation algorithm that will be described in section 5.

4. DISPARITY ESTIMATION

In this section we present a new disparity estimation scheme specially suited for the chain map.

In [2] a set of promising disparity estimation algorithms is discussed based on dynamic programming. Figure 5a shows the basic operation in these algorithms. For each point in the matching space three possible links to different disparity paths are considered. The decision is based on choosing the path with minimum cost.

We propose a new basic operation shown in Figure 5b. Only two predecessors are taken into account. The two steps correspond to chain map step bits. Locally, at the new node, a decision is made if the current left and
right pixels should be matched resulting in the chain map match bits. To make the decision we apply a threshold on the luminance difference between the left and right pixels. The decisions for step and match bits are separated in this algorithm and thus can be done in parallel.

The output of the proposed algorithm fits exactly to the chain map format and the hardware complexity is lower compared with the conventional algorithm.

5. IMAGE INTERPOLATION

In this section we compare the hardware complexity of image interpolation schemes based on disparity fields, the chain map and the reduced chain map.

To obtain the intermediate pixel coordinates, scaling of disparity vectors and additions are necessary in disparity field based interpolation algorithms [6]. Using the chain map or reduced chain map, only additions and increments are needed. The sequential nature of the chain map allows for the use of FIFO memory without addressing.

After the correct left and right pixels have been localised, the intermediate pixel luminance is defined by a weighted average of the left and right image pixels:

\[ I_M = W_L \cdot I_L + W_R \cdot I_R \]  
\[ W_L + W_R = 1 \]  

(1)  
(2)

In disparity field and chain map based algorithms, the weights basically can have three values [6]. In left occlusions \( W_R = 0 \), in right occlusions \( W_R = 1 \) and in object areas \( W_R = 1/2 + P \), with \( P \) equal to the position of the virtual intermediate camera:

\[ \text{LEFT) } -1/2 \leq P \leq 1/2 \quad \text{ (RIGHT) } \]  

(3)

In [5] we introduced a special interpolation algorithm that generates similar interpolation weights using the reduced chain map, based on step bits only. If we denote the step bits by \( S(i) \) and assign value +1 to an L stepbit and value -1 to an R stepbit, then the interpolation weights at match position \( x \) are defined by:

\[ W_R(x) = W + \Delta W \cdot ^\wedge S(x) \]  
\[ W = \frac{1}{2} + P \quad \Delta W = \frac{1}{2} \cdot |P| \]  

(4)  
(5)

The \( ^\wedge S(x) \) is equal to 0 at normal object areas, while at left and right occlusions it is 1 and -1 respectively. In these cases the weights are equal to those of common interpolation algorithms. At left contracted objects \( 0 < ^\wedge S < 1 \) and at right contracted objects \( -1 < ^\wedge S < 0 \). Here the new interpolation scheme adapts the weights in a continuous way.

The parameter \( L_w \) determines the speed with which the weights can vary. A very small \( L_w \) gives fast varying, noisy weights. A very large \( L_w \) results in almost unchanging weights. An optimum should be determined experimentally.

The weights defined by (4), (5) and (6) can be determined in a recursive way resulting in a lower hardware complexity:

\[ W_R(x+1) = W_R(x) + \frac{\Delta W}{L_w} \{ S(x+1 + ^\wedge L_w) - S(x - ^\wedge L_w) \} \]  

(7)

The expression between accolades can only have values -2, 0 and 2, so there are only three possible updates for the weight.

6. FILTERING IN THE REDUCED CHAIN MAP DOMAIN

Linear filtering is easily possible in the reduced chain map domain. The main difference with the vector field is that the implementation of the filter is memoryless in the vector field domain while it needs memory in the reduced chain map domain.

In the PANORAMA hardware system we use vertical linear interpolation on the reduced chain map, since the used block based disparity estimator outputs only disparity values at each fourth scanline while the interpolator needs dense disparity data. Interpolation of scanline \( I \) between two scanlines \( A \) and \( B \) at any intermediate position \( 0 \leq \alpha \leq 1 \) (B) can be achieved by successive approximation using only interpolations with \( \alpha = 0.5 \). First we interpolate between \( A \) and \( B \), and then re-interpolate between \( A \) or \( B \) and the midway scanline, etc. Figure 6 shows the state machine of interpolation at \( \alpha = 0.5 \). There are only two states, so the filter is very easily implemented in hardware.

Figure 6: State machine for scanline interpolation in the reduced chain map (stepbits scanline \( A \) and \( B \))
In [5] we gave an example of non linear filtering for lossy coding of the reduced chain map. In general, loss is introduced by permuting L and R step bits. The permutation process in the coding scheme is of relative low complexity compared to the other parts in the algorithm.

Not all linear filters can easily be implemented in the chain map domain. Median filters or topological filters should be applied in the disparity vector field domain and thus need conversion processes. The conversion is a simple process, implicitly described by Figure 4.

7. EXPERIMENTAL RESULTS

We generated chain maps of the stereo image sequence MAN, shown in Figure 7. Figure 8 shows the chain map obtained by the disparity estimation algorithm described in section 4. The match/step bits are indicated by white (M/L) and black (akhir).

![Figure 7: Original left and right MAN sequence](image)

![Figure 8: Chain map a) match bits b) step bits](image)

We constructed intermediate views with the algorithms of section 5 using the chain map and the reduced chain map. We found that the results look best for values of \( L_w \) between 4 and 20. Figure 9 shows the interpolated centre view based on the reduced chain map algorithm with \( L_w = 8 \). It was subjectively evaluated as a natural and high quality intermediate view. The intermediate view based on interpolation with the chain map looks almost the same. Only at the transitions between the background and the person small luminance jumps can be noticed.

![Figure 9: Synthetic centre view based on interpolation with reduced chain map](image)

Based on the two types of intermediate views and the reduced chain maps, we used image extrapolation [4] to reconstruct the original stereo image sequence.

Subjectively, the reconstructed stereo image pairs are only slightly different from the original stereo pair. Objectively, using the chain map intermediate view an average PSNR value of 41.5 dB is obtained, using the reduced chain map 43.1 dB is obtained. The improvement in PSNR is due to the contraction adaptive weights used in the interpolation algorithm based on the reduced chain map.

9. CONCLUSIONS

We presented a new disparity map format, the chain map, as an alternative to the existing vector field format. Advantages of the chain map are: one single bidirectional map in stead of two unidirectional vector fields, explicit indication of occlusions, no upper or lower bound on disparity values, no disparity offset, format takes only 4 bits per pixel.

We have presented disparity estimation and image interpolation algorithms with low hardware complexity, based on the chain map. A reduced version of the chain map with only 2 bits per pixel was introduced that further reduces the complexity of the disparity estimation algorithm.

High quality intermediate views were obtained by the disparity estimation and interpolation algorithm. Stereo image pairs were reconstructed and subjectively evaluated as only slightly different from the original image pair. Objectively, the reconstruction resembled the original with a PSNR value of 43 dB based on interpolation with the reduced chain map.

We conclude that the chain map is an efficient disparity map representation, allows for very good interpolation results and lowers the hardware system complexity substantially.

REFERENCES

[1] ACTS PANORAMA project, Research Project funded by the European Community, http://www.tnt.uni-hannover.de/project/eu/panorama


