Fast disparity estimation algorithm for mesh-based stereo image/video compression with two-stage hybrid approach

Shao-Yi Chien, Shu-Han Yu, Li-Fu Ding, Yun-Nien Huang, and Liang-Gee Chen

DSP/IC Design Lab,
Department of Electrical Engineering and Graduated Institute of Electronics Engineering,
National Taiwan University,
No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan

ABSTRACT

Disparity estimation is a very important operation in stereo image and video compression. However, existing disparity estimation algorithms require too large computation power. In this paper, we propose a fast disparity estimation algorithm for mesh-based stereo image and video compression with two-stage hybrid approach, which is names as Two Stage Iterative Block and Octagonal Matching algorithm (TS-IBOM). The first stage of this algorithm is Iterative Block Matching algorithm, which can give a good initial guess of the disparity vectors with a little computation. In the second stage, Iterative Octagonal Matching algorithm is employed to refine the disparity vectors. A local updating scheme is also proposed to further accelerate the process by skipping nodes whose disparity vectors have been derived already. Experimental results show that the proposed algorithm can give good results and is 18 times faster than Octagonal Matching. This algorithm is suitable to be employed in mesh-based stereo image and video compression systems.

Keywords: Disparity estimation, TS-IBOM, Two Stage Iterative Block and Octagonal Matching algorithm, Octagonal Matching, stereo image and video compression.

1. INTRODUCTION

Stereo image and video make people to sense three-dimensional perception by showing two images or frames captured with two cameras to each eye simultaneously. Compared with traditional mono-view video, stereo video can give users more vivid perception and can be used for 3D-TV, telepresense, and immersive communication. In stereoscopic image and video systems, disparity estimation, which can find the correspondences between the stereo pairs, is the key operation. It can give the information for inter-view prediction in compression systems. Furthermore, with the information of camera configurations and the disparity vectors, the three-dimensional models can be reconstructed.

Many disparity estimation algorithms have been proposed, such as feature based algorithms, block-matching, pel-recursive, optical flow, and Bayesian approach.[1] The adaptive window stereo matching algorithm[2] can be viewed as one of the block-matching based approach. The block size is adjusted according to the intensity and disparity variation of the two frames. It is only suitable for objects which are far from the cameras because of the frontoparallel plane assumption. Dynamic program[3] is another popular approach. It searches the matching pairs along the two corresponding epipolar lines of the two images. The computation of this kind of algorithms are large.

All the disparity estimation algorithms above are designed to derive dense disparity field. However, for video and image compression systems, coarse field are preferred since the large amount information of dense disparity field dramatically reduce the coding efficiency. Block-based coarse motion field is used for traditional video compression systems; however, it cannot be used for stereo image pairs compression since the frontoparallel plane model no longer establishes.

Further author information: (Send correspondence to Shao-Yi Chien)
Shao-Yi Chien: E-mail: shaoyi@video.ee.ntu.edu.tw, Telephone: +886 2 2363 5251 ext 332
Liang-Gee Chen: E-mail: lgchen@cc.ee.ntu.edu.tw, Telephone: +886 2 2363 5251 ext 344

Visual Communications and Image Processing 2003, Touradj Ebrahimi, Thomas Sikora,
Editors, Proceedings of SPIE Vol. 5150 (2003) © 2003 SPIE · 0277-786X/03/$15.00

On the other hand, mesh-based video coding is a new technique for motion estimation and compensation.[4] The advantage of mesh-based processing is that it can always provide better performance for both spatial and temporal interpolation. Therefore, it has been found being more effective to be used for inter-view prediction since it can provide a coarse representation of the disparity field with adequate models degenerated from three-dimensional mesh.[5]

For mesh-based disparity estimation and compensation, Hexagonal Matching algorithm (or Octagonal Matching for rectangular patches)[6] can give near-optimal solutions and is widely used. Although it can provide good performance, the computation intensity is enormous. A fast algorithm is proposed by Wangs with gradient-descent technique[5]; however, it is easy to be trapped in local minimum, and the computation intensity is still too large. Consequently, a fast algorithm is required for mesh-based disparity estimation for a stereo image and video compression system.

In this paper, we propose a novel fast disparity estimation algorithm with two-stage hybrid approach named as Two Stage Iterative Block and Octagonal Matching (TS-IBOM) algorithm. In the first stage, Iterative Block Matching algorithm can give a good initial value of the disparity vector of each node. Next, in the second stage, Iterative Octagonal Matching algorithm is applied to refine the disparity vectors. Furthermore, a local updating scheme is also proposed to further accelerate the process.

This paper is organized as follows. In Section 2, the concepts of disparity estimation for mesh-based disparity compensation are introduced. Then the proposed algorithm is described in Section 3, and the experimental results are shown in Section 4. Finally, Section 5 concludes this paper.

2. DISPARITY ESTIMATION FOR MESH-BASED DISPARITY COMPENSATION

In this section, the concepts and models of mesh-based disparity compensation are first introduced in Section 2.1. Next, in Section 2.2, the constraints of disparity vectors are described. The classical disparity estimation algorithm, Octagonal Matching algorithm, is then introduced in Section 2.3. It will be an important benchmark algorithm in this paper.

2.1. Mesh-Based Disparity Compensation

The model of mesh-based disparity compensation can be shown as Fig. 1. A physical object can always be represented with three-dimensional mesh models, that is, the surfaces of the object are modeled as several continuous small patches, where each patch is a small plane. The three-dimensional mesh models are projected to the image plane to form two-dimensional mesh. When two cameras are considered, there are two image planes, and the meshes projected on the left image plane and the right image plane are illustrated in Fig. 1. Since each patch is a small plane, the two coordinates corresponding to the same physical object can be transformed to each other with perspective or affine transform. For example, if the projected coordinate of the left eye of the man is $\mathbf{P}_l$ in the left image plane and is $\mathbf{P}_r$ in the right image plane. We can always find an affine transform W, with which $\mathbf{P}_r = W(\mathbf{P}_l)$ and $\mathbf{P}_l = W^{-1}(\mathbf{P}_r)$.

There are several assumptions for the two-dimensional mesh model. First, there is no disparity discontinuity, that is, the surface of the object is continuous. Second, no occlusion occurs. With these two assumptions, the two-dimensional mesh model can provide the same performance as that of the three-dimensional mesh model.

There are two kinds of two-dimensional mesh: regular mesh and arbitrary mesh.[4] With arbitrary mesh, the delicate surface structure can be represented with small patches, and the smooth surface can be represented with large patches. Therefore, the performance of arbitrary mesh is better. But it is not suitable to be used in a compression system since the positions of the nodes should also be transmitted to the decoder, which introduces large bit-rate overhead. Moreover, the geometry of the mesh should be reconstructed with Delaunay triangulation operation in the decoder, which introduces large computation power overhead. Consequently, regular mesh, where all the surfaces are modeled with rectangular patches in the same size, is more suitable for compression systems and is considered in this paper since only few parameters are required to represented the geometry of the mesh.

The concept of mesh-based disparity compensation is shown in Fig. 2, where the one image of the stereo pair is reconstructed (compensated) with the other image by warping the two-dimensional mesh. In this paper,

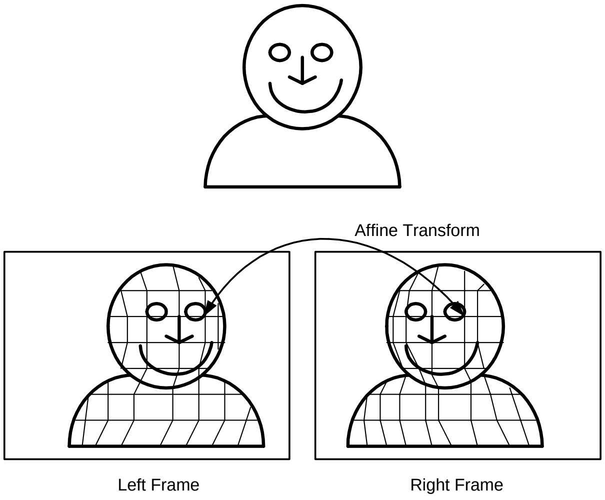

Figure 1. Projection a three-dimensional mesh to two-dimensional mesh.

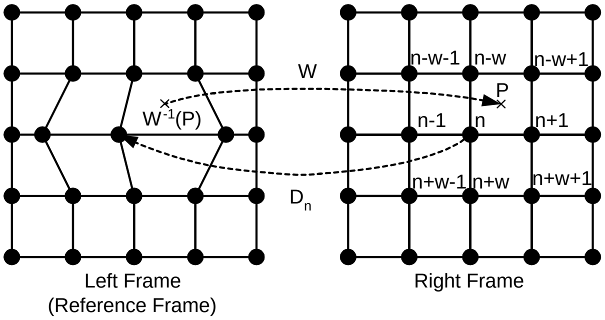

Figure 2. Illustration of mesh-based disparity compensation.

the left frame is used as the reference frame to predict the right frame. Because regular mesh is employed here, the right frame (current frame) is divided into blocks, and nodal nodes are placed at the four corners of these blocks. The texture of left frame is warped to the right frames according to the positions of the nodal nodes by use of the following equations:

for every pixel $\mathbf{P} = (x, y)$ in the right frame,

$$I'_r(\mathbf{P}) = I_l(W^{-1}(\mathbf{P})), \tag{1}$$

$$W^{-1}(\mathbf{P}) = \sum_{i \in N(\mathbf{P})} \phi_i(\mathbf{P})\mathbf{D}_i + \mathbf{P}, \tag{2}$$

where $I_r(\mathbf{P})$ is the right frame, $I_l(\mathbf{P})$ is the right frame, $I'_r(\mathbf{P})$ is the prediction frame for the right frame from the left frame, $W(.)$ is the warping function, ϕ_i is the weighting function, $\mathbf{D}_i$ is the disparity vector of node n, and $N(\mathbf{P})$ is the set of neighboring nodal nodes of the pixel $\mathbf{P}$. For example, $N(\mathbf{P}) = \{n-w, n-w+1, n, n+1\}$ in Fig. 2, where w is the number of horizontal nodal nodes. Note that bi-linear interpolation technique is applied here instead of affine transform because of its better performance.[5]

2.2. Constraints of Disparity Vectors

The disparity vectors should meet two kinds of constraints: epipolar constraint and ordering constraint.[7] To meet the epipolar constraint, the corresponding pixels in a stereo pair always lie on the respective epipolar lines. For a parallel camera configuration, namely, the two cameras are set in parallel, the epipolar lines are all horizontal. Only disparity vectors in x-direction should be considered, and the y component of the vectors are all zero. This property can dramatically simplify the computation; therefore, in this paper, we only consider parallel camera configuration. For stereo pairs captured with other camera configurations, the epipolar geometry should be first derived, and the images can be calibrated to parallel camera configuration.[8]

The second constraint is the ordering constraint, which can be expressed as the following equation.

$$\text{If } \mathbf{P}_i < \mathbf{P}_j \Rightarrow W^{-1}(\mathbf{P}_i) < W^{-1}(\mathbf{P}_j). \tag{3}$$

This constraint imply that the mesh should not cross to each other.

2.3. Disparity Estimation with Octagonal Matching

Octagonal Matching[6] is a classical approach to derive motion vectors for mesh-based systems. It can give a near-optimal solution and is widely used. The disparity vectors of the nodal nodes are derived iteratively. In each iteration, the disparity vectors are decided in raster scan manner. For each node n,

$$D_n = \arg\min_{D_n} \sum_{\mathbf{P}\in NB(n)} |I_r(\mathbf{P}) - I'_r(\mathbf{P})|, \tag{4}$$

where

$$x_{n-1} + D_{n-1} \leq x_n + D_n \leq x_{n+1} + D_{n+1}, \tag{5}$$

$NB(n)$ is the neighboring four blocks of node n, which are shown by gray color in Fig. 3, and x_n is the x component of the position of node n. To find the disparity vector, the iterative full-search strategy is employed to find the corresponding position $W^{-1}(\mathbf{P}_n)$ of the nodal node $\mathbf{P}_n$. Only the disparity vector of one node is considered at one time, where the other disparity vectors are fixed. For each candidate disparity vector in the range derived from (5),

$$x_{n-1} - x_n + D_{n-1} \leq D_n \leq x_{n+1} - x_n + D_{n+1}, \tag{6}$$

the four transforms of the four blocks in $NB(n)$ are calculated and the texture of $NB(n)$ in the left image is warped to the right image as shown with the gray regions in Fig. 3, and the sum of absolute difference between the warped texture and the original texture is calculated. After all the sum of absolute difference values of all the candidates are calculated, the candidate with the minimum value is the disparity vector D_n of this node n, as shown in (4). Next, the disparity vector D_{n+1} is calculated. After all the disparity vectors are derived, the next iteration starts, where the same process is applied from the first node to the last node. This procedure is applied iteratively until no change of disparity vectors occur. Note that for parallel camera configuration, the vector $\mathbf{D}_n$ is degenerated to scalar D_n, which is the x component of the disparity vector. Equation (5) is just the ordering constraint for the object surface continuity assumption of mesh.

Although Octagonal Matching can provide good performance, the computation intensity is enormous because of the full search strategy. A fast algorithm is proposed by Wangs[5] with gradient-descent approach; however, it is easy to be trapped in local minimum, and the computation intensity is still too large.

3. PROPOSED ALGORITHM

In this section, we propose a new fast algorithm for mesh-based disparity estimation named as Two Stage Iterative Block and Octagonal Matching (TS-IBOM) algorithm. This algorithm contains two stages. The first stage is Iterative Block Matching algorithm, which can derive good initial values of disparity vectors quickly. Next, in the second stage, the Iterative Octagonal Matching algorithm can further refine the disparity vectors. The local updating and shape-adaptive schemes are also proposed to further accelerate the process.

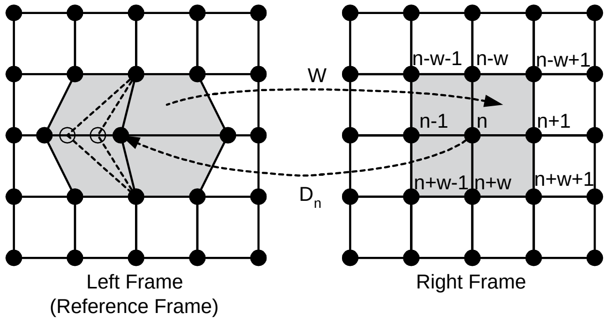

Figure 3. Illustration of Octagonal Matching algorithm.

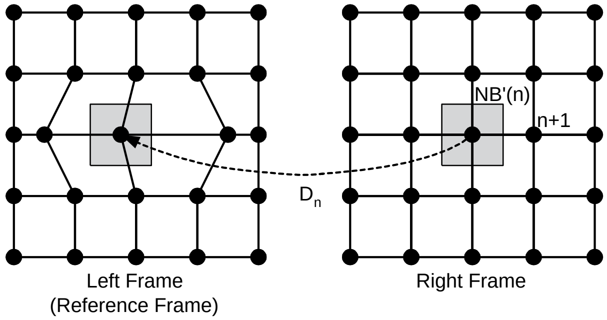

Figure 4. Illustration of Iterative Block Matching algorithm.

3.1. Stage One: Iterative Block Matching Algorithm.

In the first stage, Iterative Block Matching algorithm is proposed. In each iteration, for each node n, we fixed the disparity vectors of the other nodes. The disparity vector can be found with the following equation:

$$D_n = \arg\min_{D_n} \sum_{\mathbf{P} \in NB'(n)} |I_r(\mathbf{P}) - I_l(\mathbf{P} + \mathbf{D}_n)|, \tag{7}$$

where $NB'(n)$ is a block centered at the node n in the right image, as shown in Fig. 4, $\mathbf{D}_n$ is a disparity vector with x component is D_n and y component is 0, and the candidates of disparity vectors are also the same as (6). The same procedure is repeated iteratively until the disparity vectors converge. Since no warping operations are involved in the procedure, block matching algorithm is much faster than Octagonal Matching. On the other hand, the performance is better than the traditional block matching algorithm because the search range is unlimited and the ordering constraint can be maintained as well with the iterative approach. Therefore, the first stage can derive a good initial guess for the disparity vectors with a little computation; however, it may be failed in some portions whose disparity vector field are not smooth, such as the boundaries and the noses of the objects, which is due to the model mismatch between block matching and stereo video.

3.2. Stage Two: Iterative Octagonal Matching algorithm

In the second stage, the Iterative Octagonal Matching algorithm, which is similar to the one-step search algorithm,[9] is employed to refine the disparity vectors. For every node in each iteration,

$$D_n = \arg\min_{D_n} \sum_{\substack{\mathbf{P} \in NB(n) \\ (D_n \mod 2^m)=0}} |I_r(\mathbf{P}) - I'_r(\mathbf{P})|, \tag{8}$$

where m is set to 2 in the first iteration, set to 1 in the second iteration, and set to 0 in the following iterations. Similarly, the procedure stops after the disparity vectors converge. The main concept of this algorithm is

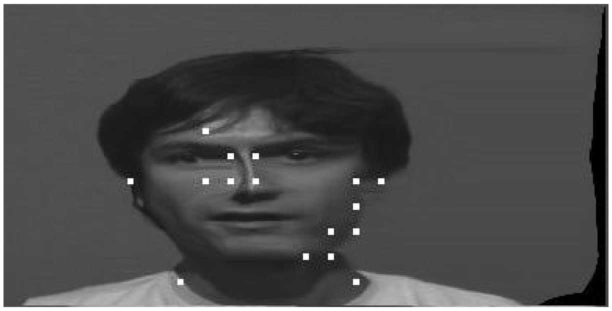

Figure 5. Nodal nodes whose disparity vectors are required to be further refined after five iterations in stage one.

described as follows. The computation of Octagonal Matching algorithm in the first several iterations is redundant since the derived disparity vectors are invalid when the disparity vectors of the neighbored nodal nodes are uncertain. Compared with Octagonal Matching, the proposed algorithm is a coarse-to-fine approach, where a coarse results can be found in the first two iterations, and the fine results can be derived in the last several iterations. The redundant computation of the first several iterations are removed with coarse search strategy. On the other hand, the accuracy of Octagonal Matching is still maintained with the fine search strategy in the last several iterations. Consequently, the Iterative Octagonal Matching algorithm can perform as well as the Octagonal Matching algorithm with fewer computation power.

3.3. Local Updating and Shape-Adaptive Scheme

In the experiments, we found that after several iterations, most of the disparity vectors are derived, and only a few nodes need to be refined. To further accelerate the process, we also propose a local updating scheme, where disparity estimation is only applied on the nodes within regions having large prediction error. For every node in each iteration, the difference of the disparity compensated frame and the right image of the four blocks connected to the current node is evaluated. If the difference is less than a threshold, the current node will not need to be processed in the next iteration and can be skipped. This scheme can accelerate the process with negligible quality degradation. An example is shown in Fig. 5. After five iterations in stage one, the nodal nodes which need to be refined are marked. It shows that only the nodes near the nose and the boundaries are marked. The disparity vectors of these nodes are then further refined with Iterative Octagonal Matching algorithm.

On the other hand, the assumptions of mesh-based disparity compensation imply the disparity estimation algorithm should be applied on each video object separately, that is, a video object segmentation algorithm[10] may be required as a pre-processing. If the video object masks are available, a shape-adaptive scheme can skip nodal nodes which do not belong to the video object.

4. EXPERIMENTAL RESULTS

The test platform is a PC with a Pentium-4 1.6 GHz processor. The standard stereo test sequence Anne, whose size is 720x288, is taken as an example in Fig. 6. The block size is 16x16. Figure 6(a) is the left frame, and Fig. 6(b) is the right frame. The target of the proposed algorithm is to generate a good prediction frame for Fig. 6(b) only from the information of Fig. 6(a). After the first stage, the result is shown in Fig. 6(c). It is obvious that the disparity vectors of the nodes near the boundaries are not correct. After the process of the second stage, the result is shown in Fig. 6(d), which is very similar to Fig. 6(b). It is shown that with mesh-based disparity compensation and the proposed disparity estimation algorithm, the left frame can give good prediction for the right frame, which is very valuable in a stereo coding system. The corresponding two-dimensional meshes are shown in Fig. 6(e) and Fig. 6(f). Note that the regular mesh is employed here, which is shown in Fig. 6(f).

Similar results can be found in Fig. 7, where the other sequence Man sizing in 386x193 is tested. After the process of the first stage, the compensated frame is shown in Fig. 7(c), where the compensated regions near the

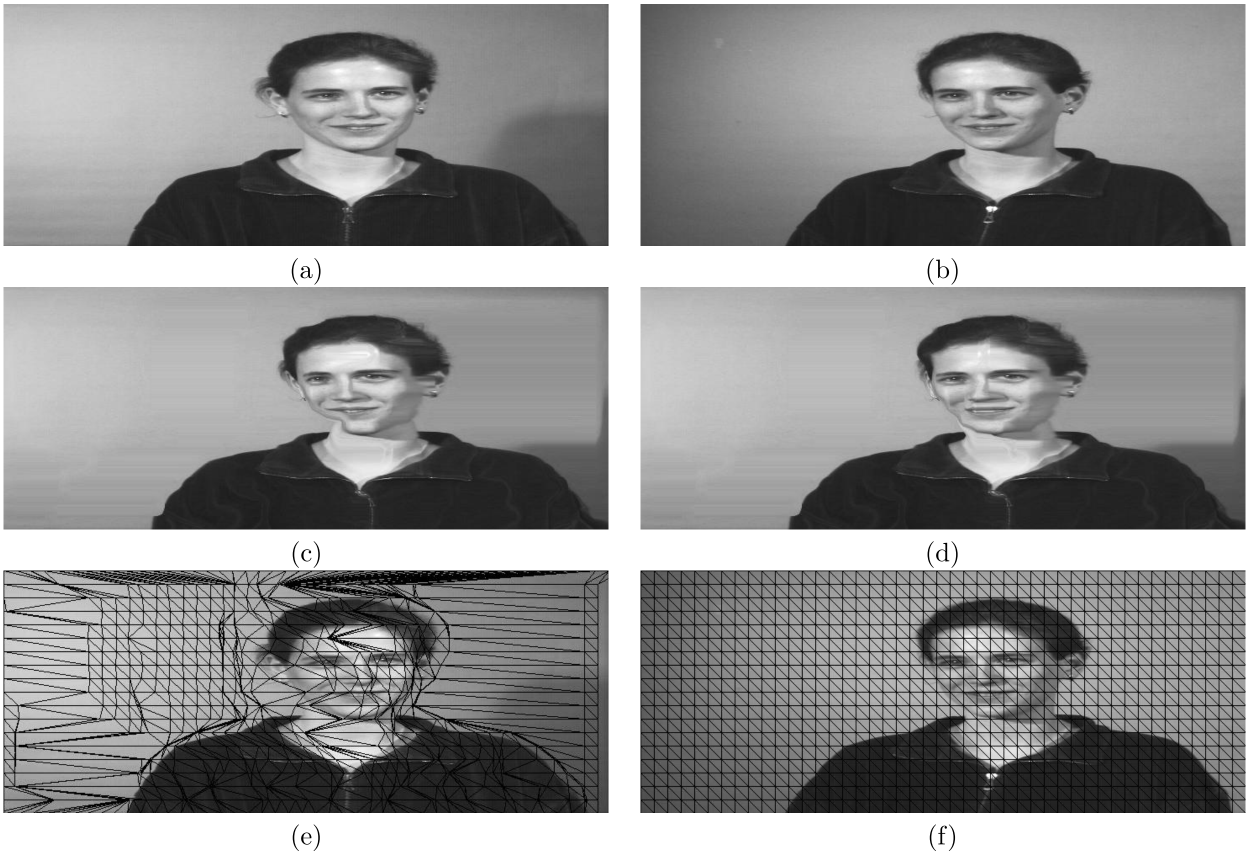

Figure 6. (a) Origin left frame of Anne #001. (b) Origin right frame of Anne #001. (c) Warped frame after the first stage. (d) Warped frame after the second stage. (e) The corresponding mesh of the left frame. (f) The regular mesh of the right frame.

nose and the boundaries of the face are not correct. After the second stage, the result is shown in Fig. 7(d). It is very similar to Fig. 7(b).

The subjective comparison of each iteration of Octagonal Matching, TS-IBOM, and TS-IBOM with local updating scheme (TS-IBOM+LU) is shown in Fig 8. It shows the disparity vectors of the proposed algorithms converge more quickly than those of Octagonal Matching. The final results of TS-IBOM are very similar to that of Octagonal Matching with less computation power. Therefore, it is more effective than Octagonal Matching. Compared TS-IBOM with TS-IBOM+LU, it shows that the performance of these two algorithms are similar. It is proven that the local updating scheme can further accelerate the disparity estimation process with negligible quality degradation. The objective comparison is shown in Fig. 9, where only the foreground object is considered, and similar results can be found. Octagonal Matching can achieve 30.51 dB in PSNR with nine iterations, TS-IBOM can achieve 30.10 dB with seven iterations, and TS-IBOM+LU can achieve 28.65 dB with seven iterations, too. Note that, although the PSNR is 1.86 dB lower than that of Octagonal Matching, the subjective quality is almost the same, as shown in Fig. 8.

The PSNR-Runtime chart is shown in Fig. 10. To achieve similar subjective quality, the runtime of Octagonal Matching, TS-IBOM and TS-IBOM+LU are 45,460 ms, 8,230 ms, and 2,420 ms. It shows that TS-IBOM is 5.52 times faster than octagonal, and TS-IBOM+LU is 18.79 faster. Note that the gradient-descent based fast algorithm[5] is only twelve times faster. Figure 10 also shows that the proposed two algorithms can achieve high PSNR with only short runtime, which are more effective than Octagonal Matching.

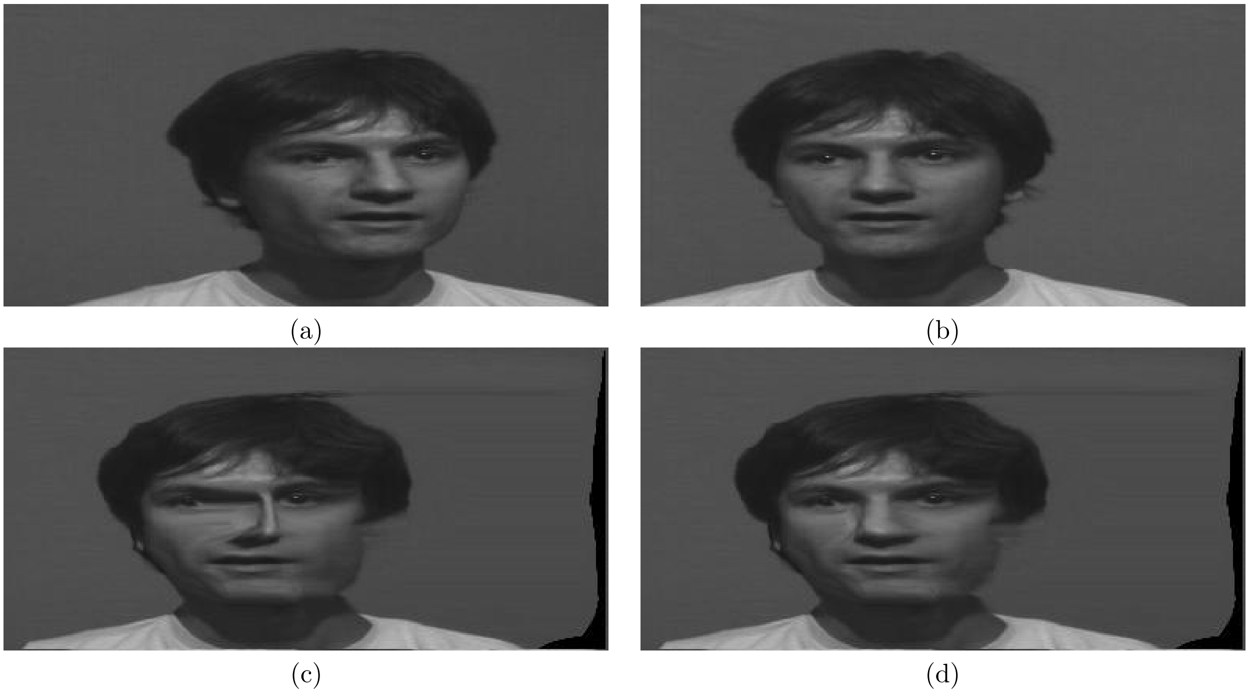

Figure 7. (a) Origin left frame of Man #001. (b) Origin right frame of Man #001. (c) Warped frame after the first stage. (d) Warped frame after the second stage.

5. CONCLUSION

In this paper, we propose a two-stage fast disparity estimation algorithm for mesh-based image and video compression named as TS-IBOM. With Iterative Block Matching algorithm and Iterative Octagonal Matching algorithm, the proposed algorithm can give good prediction frame with less computation. It can be further accelerated with local updating scheme, by which 18 times speedup can be achieved compared with Octagonal Matching.

REFERENCES

1. A. Redert, E. Hendriks, and J. Biemond, "Correspondence estimation in image pairs," *IEEE Signal Processing Magazine* **16**, pp. 29–46, May 1999.
2. T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**, pp. 920–932, Sept. 1994.
3. N. Grammalidis and M. G. Strintzis, "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," *IEEE Transactions on Circuits and Systems for Video Technology* **8**, pp. 328–344, June 1998.
4. A. M. Tekalp, P. V. Beek, C. Toklu, and B. Gunsel, "Two-dimensional mesh-based visual-object representation for interactive synthetic/natural digital video," *Proceedings of the IEEE* **86**, pp. 1029–1051, June 1998.
5. R.-S. Wang and Y. Wang, "Multiview video sequence analysis, compression, and virtual viewpoint synthesis," *IEEE Transactions on Circuits and Systems for Video Technology* **10**, pp. 397–410, Apr. 2000.
6. Y. Nakaya and H. Harashima, "Motion compensation based on spatial transformations," *IEEE Transations on Circuits and Systems for Video Technology* **4**, pp. 339–356, June 1994.
7. Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*, Prentice Hall, New Jersey, 2002.

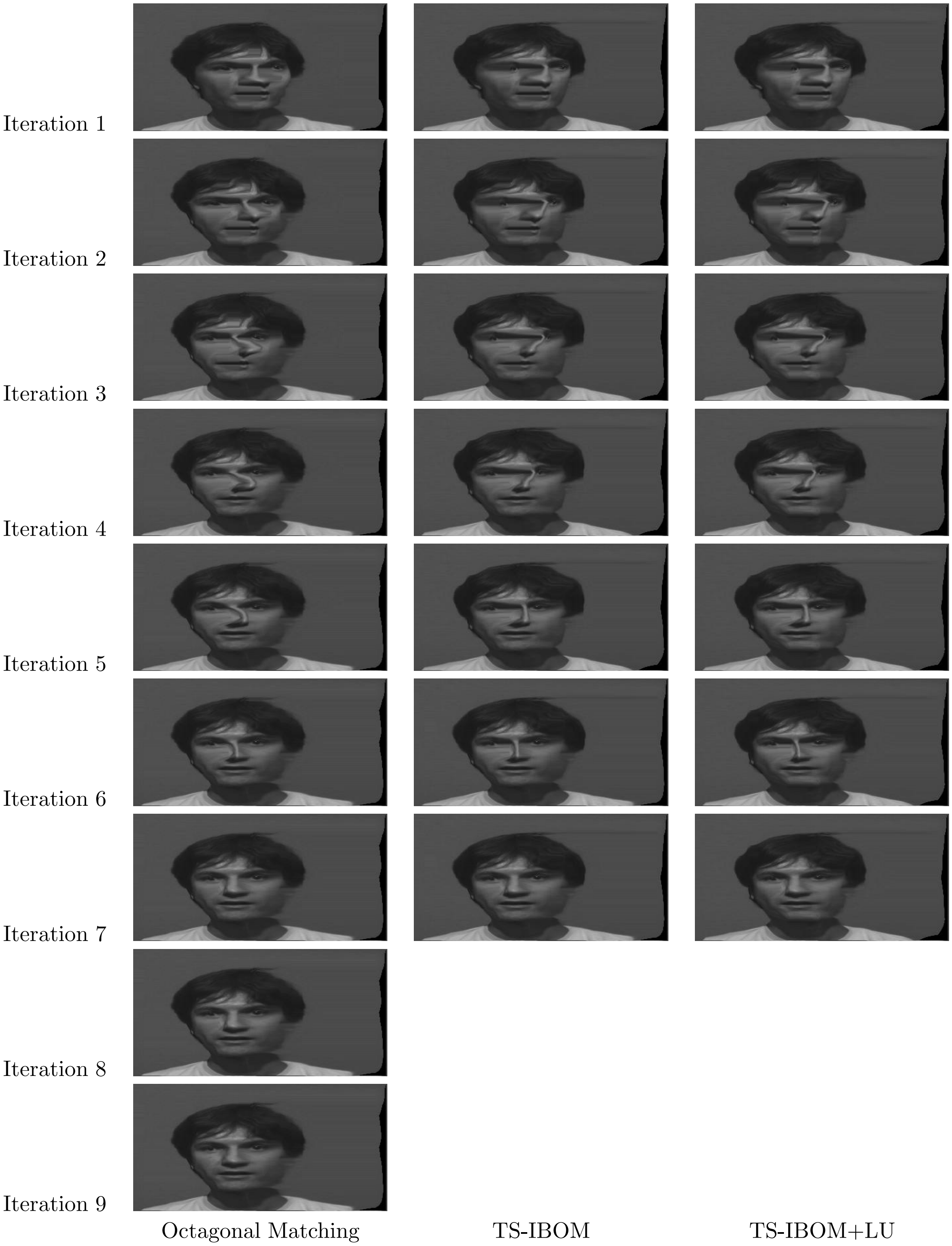

Figure 8. Subjective comparison of each iteration of Octagonal Matching, TS-IBOM and TS-IBOM+LU.

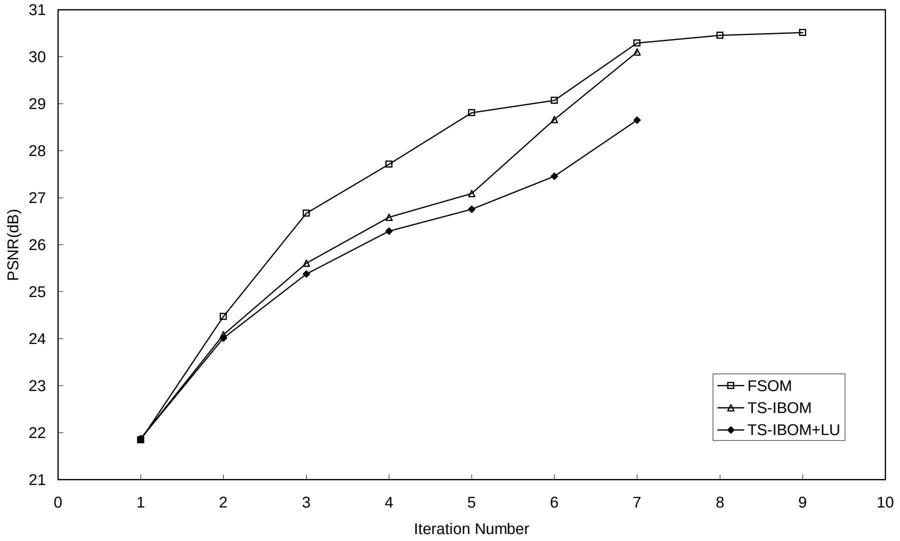

Figure 9. PSNR-Iteration chart.

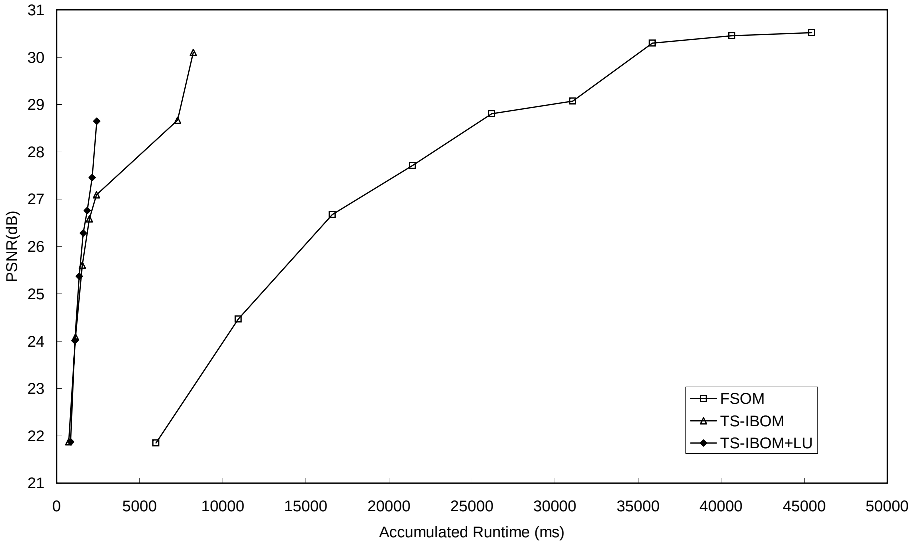

Figure 10. PSNR-Runtime chart.

8. Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Technical Rerport No. 2927, INRIA* , 1996.
9. G. Heising, "Efficient and robust motion estimation in grid-based hybrid video coding schemes," in *Proceedings of International Conference on Image Processing 2002*, pp. 687–700, 2002.
10. S.-Y. Chien, S.-Y. Ma, and L.-G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Transactions on Circuits and Systems for Video Technology* **12**, pp. 577–586, July 2002.