Control Points Based Semi-Dense Matching

Yunde Zhong, Huaifeng Zhang
National Laboratory of Pattern Recognition
Institute of Automation
Chinese Academy of Sciences 100080 P.R.China
{ydzhong, hfzhang } @nlpr.ia.ac.cn

Abstract

The common practice to get more matching points between two perspective images proceeds from coarse matching to dense matching. However, in some applications, for example, in the visualization of point-based 3D reconstruction results, one usually needs more matching points than those which could be obtained in coarse matching process to reveal sufficient structures, and does not need clouded matching points from dense matching process which is additionally of heavy computational load. In addition, it is sometimes desirable that the matching points could be evenly distributed across the whole image. The technique proposed in this note is a tentative step to this end. By adaptively adjusting the related parameters in the propagation stage, an appropriate number of matching points are obtained and distributed evenly. The extensive experiments validate our proposed new technique.

Key words: *stereo vision, dense matching, sparse matching, visualization of 3D structure*

1. Introduction

It is well known that the stereo matching is of a crucial problem in computer vision. Matching as many as possible points and avoiding as many false matches as possible are two aspects of the stereo matching. From these two aspects, the methods of stereo matching are roughly classified into two categories: feature-based one and area-based one.

Feature-based methods extract prominent features, such as corners and lines, and match them across two or more views [5][6][12]. They yield only sparse disparity maps. Area-based methods yield a dense disparity map by matching small image patches as a whole, and the underlying assumption appears to be valid for textured areas in image pairs with small difference [1][3][7][8][9]. A comparison of stereo matching can be found in [10].

In a 3D reconstruction system, sparsely matched points could only stretch out the coarse outline of 3D objects but lacking of sufficient details. On the other hand, although dense matching is a theoretically perfect scheme, in practice, it is usually inevitable to obtain false matches which are fatal to 3D reconstruction. Does there exist an intermediate state that can depict sufficient details but avoid many false matching points? Motivated by this idea, we propose a new and efficient method called semi-dense matching in this note. It is in fact a tradeoff between the sparse matching and the dense matching.

Our technique is based on some highly reliable matches called control points derived from sparse matching. For each pair of control points, a reliable degree is evaluated to construct a propagating model. In each propagating area, an expectant number of corners are extracted using adjusted Harris operator. According to the optimum principle and the adaptive principle, which will be discussed later, more new corresponding points are found. In the end, we obtain an expectant number of matching points which distribute evenly across image.

The paper is organized as follows: in section 2, we describe the control points and the propagation model which is of primary importance. Then we discuss the implementary details in section 3. Extensive experiments are reported in section 4. Finally some conclusions and discussions are provided in section 5.

2. Control Points and Propagation Model

2.1 Control Points and Their Attributes

In general, there exits disparity continuity in a pair of stereo images. In the neighborhoods of a pair of matched points, there should exit some other pairs of matching points that are to some degree related to the matched pair. In this paper, we define this pair of matched points as control points. They are representatives of local features. And the process to get new pairs of matching points is called propagation.

Each pair of control points has two important attributes, the disparity and the reliable degree. In a pair of images, *view1* and *view2* shown as in Figure 1, point p with the coordinates (u,v) in *view1* and point p' with the coordinates (u',v') in *view2* are a pair of control points. The disparity ρ is the difference of coordinates between the two control points.

$$\rho = p' - p = \begin{bmatrix} u' - u \\ v' - v \end{bmatrix}$$

A pair of control points (p, p') should satisfy the epipolar constraint and the correlation constraint within a fixed tolerance. The reliable degree ψ defined below is used to measure quantitatively the fitness of matching.

$$\psi(r,d,d') = r \cdot f\left(\sqrt{d^2 + d'^2}\right)$$

$$f(x) = \begin{cases} 1 - \dfrac{x}{\sigma} & if\ 0 \le x \le \sigma \\ 0 & else \end{cases}$$

where r is the zero-mean normalized cross-correlation measuring for 7×7 windows around the points p and p' respectively, σ is a threshold to describe the extent satisfying the epipolar constraint. d and d' are the distance of points p and p' to the corresponding epipolar lines outlined below.

If the fundamental matrix F is known, the corresponding epipolar line of p' can be expressed as :

$$l = F^T \tilde{p}' = \begin{bmatrix} f_{11} & f_{21} & f_{31} \\ f_{12} & f_{22} & f_{32} \\ f_{13} & f_{23} & f_{33} \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix}$$

And the epipolar line corresponding to p is

$$l' = F\tilde{p} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} l'_1 \\ l'_2 \\ l'_3 \end{bmatrix}$$

where $\tilde{p}$ and $\tilde{p}'$ are the homogeneous coordinates of the control points p and p', i.e.:

$$\tilde{p} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad \text{and} \quad \tilde{p}' = \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}.$$

Then d and d' can be respectively expressed as,

$$d = \frac{|ul_1 + vl_2 + l_3|}{\sqrt{l_1^2 + l_2^2}} \quad \text{and} \quad d' = \frac{|u'l'_1 + v'l'_2 + l'_3|}{\sqrt{l_1'^2 + l_2'^2}}$$

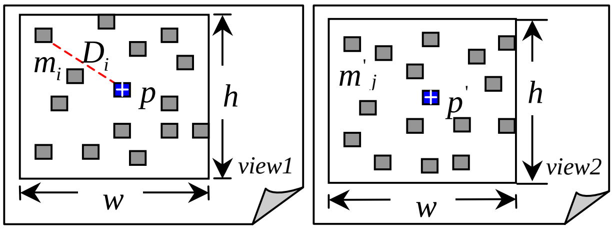


Figure 1. *Control points* (p, p') *and the propagating areas*

2.2 Propagation Model

The propagation area is a $w \times h$ window centered at the control points. The window's sizes are adjustable based on the reliable degree as:

$$w = 100 + 20 \times (\psi - 0.5)$$
$$h = 100 + 20 \times (\psi - 0.5)$$

In each propagating area, the expectant number of corner points is set in a proportion to its area. With this strategy, only prominent corners in texture-rich areas are extracted, but in areas lacking of texture, non-prominent corners are also extracted. These evenly distributing corners make it possible to obtain matching points distributed evenly across image.

Figure 1 demonstrates the implementation details. Let $m_i, m'_j, (i = 1 \cdots n, j = 1 \cdots n')$ be corner points in each propagation area. Let $M = \{m_i \mid i = 1 \cdots n\}$ and $M' = \{m'_j \mid j = 1 \cdots n'\}$ be corner sets. Potential matching points are (m_i, m'_j), $m_i \in M$, $m'_j \in M'$. If potential points (m_i, m'_j) are matching points, they should have similar disparity to that of the control points.

From the research of psychophysics [13], the relative disparity δ should have an upper limit expressed as below:

$$\delta = \left(m'_j - m_i\right) - \rho$$

$$\|\delta\| \le \frac{2K}{2-K} D_i \,,$$

where K is a coefficient relating to psychophysics, D_i denotes the distance between m_i and p.

Correlation is another important criterion in matching process. Traditionally, correlation is carried out on windows centered at two points respectively, but in our propagation process, we use a new scheme (Figure 2).

Firstly, we partition the propagating area into 4 subareas marked with I,II,III and IV. For each potential matching points $\left(m_i, m'_j\right)$, we introduce a new variable O_i to describe their orientations in the first image. $O_i = \left[O_{xi}, O_{yi}\right]^T$,

$$o_{xi} = \begin{cases} -1 & \text{if } m_i \text{ in part I} \\ 1 & \text{if } m_i \text{ in part IV} \\ 0 & \text{Other cases} \end{cases}$$

$$o_{yi} = \begin{cases} -1 & \text{if } m_i \text{ in part III} \\ 1 & \text{if } m_i \text{ in part II} \\ 0 & \text{Other cases} \end{cases}$$

Then the correlation r_{ij} is computed using the following formula:

$$r_{ij} = \frac{\sum_l \sum_k \left(p_{(x_i+l)(y_i+k)} - \overline{P}_i\right)\left(p'_{(x'_j+l)(y'_j+k)} - \overline{P}'_j\right)}{\sqrt{\left(\sum_l \sum_k \left(p_{(x_i+l)(y_i+k)} - \overline{P}_i\right)^2\right)\left(\sum_l \sum_k \left(p'_{(x'_j+l)(y'_j+k)} - \overline{P}'_j\right)^2\right)}}$$

$$\overline{P}_i = \frac{\sum_l \sum_k p_{(x_i+l)(y_i+k)}}{7 \times 7} \,,$$

$$\overline{P}'_i = \frac{\sum_l \sum_k p'_{(x'_j+l)(y'_j+k)}}{7 \times 7}$$

where l changes from $3\,o_{xi}$ to $7 + 3\,o_{xi}$, k changes from $3\,o_{yi}$ to $7 + 3\,o_{yi}$.

Experiments show that this new correlation scheme can generate more and relatively accurate matching points in a local area, especially those allocated near the edge of objects in the scene whose depth changes sharply.

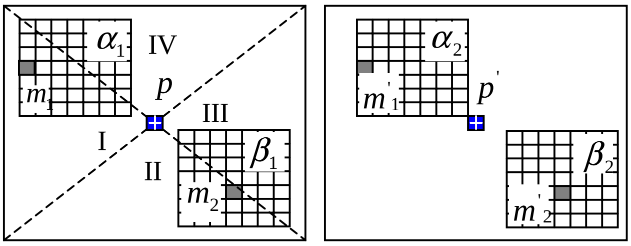


Figure2. *Using areas* α_1 *and* α_2 *to compute the correlation between the potential matching points* $\left(m_1, m'_1\right)$, β_1 *and* β_2 *for points* $\left(m_2, m'_2\right)$

3. Semi-Dense Matching Process

3.1. Principles

3.11 Optimum Principle. In every iterative process, the pair of control points with the highest reliable degree is selected to propagate first, and the less reliable one will be postponed till it becomes prominent. With this strategy, the newly matched points from the more reliable control points will inhibit the propagation of less reliable ones.

In each propagating process, a corner in one image will often have more than one candidates in the other image, or vice versa. In these cases, the pair of points with the highest correlation value is selected to avoid the ambiguity.

3.1.2. Adaptive Principle. In general, thresholds used in the propagation in the whole image are prefixed and keep unchanged. The resulting disadvantage is that abundant matching points will be obtained in texture-rich areas and few in areas lacking of texture. Naturally, it will be desirable if thresholds used in the propagation can vary according to their local features.

In our method, a propagation model is constructed in each propagating area to generate new matching points. Because the attributes of the control points are available reliable information about the local features, they are used to dynamically adjust the parameters of each propagation model to get optimum result. Thus texture-rich areas can be accurately matched and the areas with poor texture will not lose corresponding points. Eventually, the matching points will distribute evenly across image.

3.2. Implementation details

The initial control points are obtained through sparse matching. Then a propagating process is invoked to generate iteratively new matching points. In all these steps, an important data structure, a special stack, is adopted to store all pairs of points and to handle optimum selections.

Two important operations are related to this stack. One is *push* operation by which a pair of points is pushed into the stack and sorted according to their reliable degree. Another is *pop* operation by which the pair of points with the most reliable degree will be popped out.

3.2.1. Initial Control Points. In this paper, the control points are obtained by the method proposed by Zhang [12]. Corners in the two images are extracted firstly using traditional methods, then the initial matches are refined by the cross-correlation. With these initial matches, we can recover the corresponding epipolar lines. To figure out the most faithful fundamental matrix, the RANSAC based 8-points algorithm [2] [11] is used.

Even so, there may be some mismatchings. Using the disparity gradient limit in [13], the pairs with large disparity errors will be discarded. All the remaining pairs of points are considered as control points. Each pair of control points has a correlation value and two distances to the epipolar lines. We compute their reliable degree using the formula described in section 2. Then all these well-chosen control points are pushed into a global stack for later use.

3.2.2. Propagating Process. In a propagating process, it is important to decide which control points propagate first. If in arbitrary order, we will get unfavorable results. According to the forementioned optimum principle, only the most reliable control points will be privileged to propagate. So each time we retrieve a pair of control points $\left(p, p^{'}\right)$ and its reliable degree ψ by a *pop* operation on the global stack.

Based on the reliable degree, the local propagation model is constructed, and the width and height of propagating windows are determined. Applying adjusted Harris detector on these two areas, we can extract corner sets M and $M^{'}$. For every pair of points $\left(m_i, m^{'}_j\right)$, $m_i \in M, m^{'}_j \in M^{'}$, we will validate them by exerting the disparity continuity model and their correlation. Only those pairs which satisfy the disparity continuity, epipolar constraint and correlation constraint are called potential matching points and are pushed into a local stack

For each pair of potential matching points $\left(m_i, m^{'}_j\right)$, $m_i \in M, m^{'}_j \in M^{'}$, its reliable degree is computed. According to the reliable degrees, potential matching points will be ranked in the local stack.

Then all the potential matching points are popped out from the local stack one by one. If they have not been matched yet, then they will be pushed into the global stack. Iterating the propagation process until the global stack is empty, we can extend matching areas to bring out more vivid visual effects in 3D reconstruction applications.

4. Experimental Results and Applications

Our new method has been tested by many pairs of stereo images (1024×768). Here we only report three examples of semi-dense matching and its applications in 3D reconstruction due to the limited space. All the computing processes were completed in only a few minutes on a personal computer 800MHZ. In each example we provide a pair of images, the comparative results by sparse matching and semi-dense matching, the 3D reconstruction results using sparse matching points or semi-dense matching points.

The first example is shown in Figure 4. Figure 4d (left) shows that 3D reconstruction based on sparse matching points (4b) did not well sketch out the outline of 3D objects especially in the areas which are lack of texture (marked by a red ellipse). From Figure 4d(right) it is obvious that our propagating method can generate new matching points within the green box to keep its shape (4c).

A second example in Figure 5 shows that our method can distribute the matching points evenly across image. This distribution can make outstanding the details of the elephants. From the right image of Figure 5c it is evident that the trunks of two elephants have been rendered satisfactorily. Another example in figure 6 shows the vivid visual effects in 3D reconstruction using semi-dense matching.

5 Conclusions and Future Work

In this paper, a new matching technique is proposed. This new technique can obtain a fairly good number of corresponding points in both texture-rich areas and texture-poor areas. The proposed technique can be considered as an intermediate one between the coarse matching and the dense matching both widely used in the literature. The chief motivation of our work is to effectively visualize a cloud of reconstructed 3D points. Of course, its potential applications are by no means limited to such a specific field.

Although the experimental result is inspiring, there are still a lot to be improved. For example, semi-dense matching with sub-pixels will probably strengthen the accuracy of matching. And the application domain of semi-dense matching is to be extended. Its application in image interpolation or morphing technique is under investigation. How to fuse the corresponding points in a series of images is another issue to pursue further in the future.

Acknowledgements

We would like to thank Li Hua and Hu Jinhui for fruitful discussions, Meng XiaoQiao for his program on sparse matching and Fei XiuYan for her program on triangulating, texturing and displaying 3D objects. This work was supported by "973" Program (G1998030502), The National Science Foundation of China (60033010, 69975021).

References

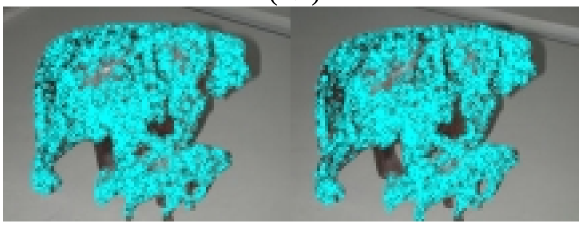
[1] C.H.Chou and Y.C.Chen, Moment preserving pattern matching, *Pattern Recognition*, 23(5): 461-474, 1990.
[2] M.A.Fischler and R.C.Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6): 381--395, 1981
[3] Pascal Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine Vision and Applications, 6(1): 35-49, 1993.
[4] A.Goshtasby, S.H.Gage and J.F.Bartholic, A two-stage cross correlation approach to template matching, IEEE Trans. *PAMI*, 6(3): 374-378, May 1984.
[5] W.E.L.Grimson. Computational experiments with a feature based stereo algorithm. IEEE Trans. *PAMI*, 7(1): 17-34, 1985.
[6] R.Horaud and T.Skordas, Stereo correspondence through feature grouping and maximal cliques, IEEE Trans. *PAMI*, 11(11): 1168-1180, 1989.
[7] M.Lhuillier and L.Quan, Image Interpolation by Joint View Triangulation, CVPR'99, Colorado, June 1999.
[8] L.Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. Proc.4th ECCV, Cambridge, UK, April 1996.
[9] .Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. International Journal of Computer Vision, 28 (2): 155-174, July 1998.
[10] R.Szeliski and R. Zabih. An experimental comparison of stereo algorithms. In *International Workshop on Vision Algorithms*, pages 1-19, Kerkyra, Greece, September 1999.
[11] P.Torr and D.Murray, The development and comparison of robust methods for estimating the fundamental matrix, *International Journal of Computer Vision*, 24(3): 271-300, 1997.
[12] Z.Zhang, Flexible camera calibration by viewing a plane from unknown orientations, in *Proc 7th ICCV*, pp. 666-673, Greece, 1999.
[13] Z.Zhang and Y.Shan, A progressive scheme for stereo matching, *ECCV Workshop SMILE2*, Dublin, June 2000.

(4.a) (4.c)

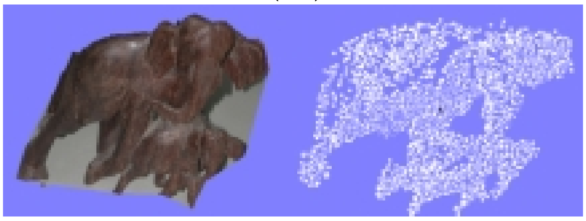
(4.b) (4.d)

Figure 4. (a) A pair of images about a green box taken form our lab. (b) the results of sparse matching (marked by blue cross). (c) the results of semi-dense matching. (d) the resulsts of 3D reconstruction from sparse matching(left) and semi-dense matching(right).

(5.a)



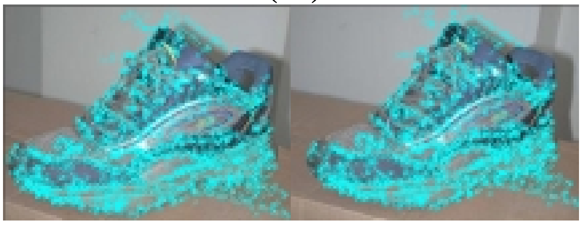
(5.b)



(5.c)

Figure 5. (a) A pair of images about two wooden elephants. (b) the results of semi-dense matching. (c) the results of 3D reconstruction(left) and the cloud of reconstructed 3D points(right)

(6.a)



(6.b)

(6.c)

Figure 6. (a) A pair of images of sneakers. (b) the result of semi-dense matching. (c) The results of 3D reconstruction(left) and the cloud of reconstructed 3D points(right)