

# DENSE CORRESPONDING PIXEL MATCHING USING A FIXED WINDOW WITH RGB INDEPENDENT INFORMATION

Teng Hu, Huayi Wu\*

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 430079 Wuhan,  
huteng@whu.edu.cn, wuhuayi@whu.edu.cn, moblemap@gmail.com.

**KEY WORDS:** Image process, Dense corresponding pixel matching, Markov Random Field, Fixed Window, RGB independent Information.

## ABSTRACT:

A fixed window based algorithm is proposed for dense corresponding pixel matching between epipolar images. The algorithm makes full use of three colour components, R, G and B, and simplifies determining the optimal window. In order to reduce the impact of obvious colour difference between corresponding pixels, the proposed algorithm uses more information from neighbouring pixels, and adjusts the contributions of the central pixels and their neighbouring pixels. Two pairs of epipolar images are tested using the fixed window based algorithm. The experimental results, including boundary testing, sampled corresponding pixels' evaluation and 3D point cloud check, demonstrate the effectiveness of the proposed algorithm.

## 1. INTRODUCTION

Corresponding pixel matching is one of the basic topics in photogrammetry. Acquiring a large quantity of accurate corresponding pixels between epipolar images remains a difficult problem. Four major interferences cause the difficulty: obvious colour differences between corresponding pixels, sensor noise, moved objects and occlusion. Generally, obvious colour differences between corresponding pixels are the primary error source for matching a pair of epipolar images. Thus, a reliable algorithm must seriously reduce the impact of this primary error source.

Current dense window-based pixel-to-pixel matching techniques contain two main types: adaptive-window and multiple-window methods. Adaptive-window methods find a proper optimal support window for each pixel. Kanade and Okutomi (1994) select an optimal window by evaluating local changes of intensity and disparity, the method depends on the initial disparity. Meanwhile, the support window is a rectangle and therefore not suitable for pixels in discontinuous areas. Boykov et al. (1998) chose an arbitrarily shaped connected window using plausibility hypothesis testing for each pixel, but the computation was very complex. Veksler(2002,2003) found a few useful windows sizes and shapes, which work well in determining support size, however the method is not generalizable. Yoon (2006) presented an adaptive support-weight approach for corresponding search; the approach can produce accurate piecewise smooth disparity maps, but a long runtime is inevitable. Multiple-window methods choose an optimal support window from several predefined windows having the same shape with different positions. Fusiello et al. (1997) assigned each pixel 9 different windows and retained as the optimal window the one window with the smallest matching cost. Kang et al. (2001) examined each window containing the pixel of interest in multi-view processing. Although these methods perform well in building relationships between matching pixels under some conditions, they have imitations: the local support window is not generalizable, requires many user-specified parameters or need long runtimes.

Based on the basic Bayesian rule, a Markov Random Field (Geman and Geman, 1984) based algorithm is formulated for

dense corresponding pixel matching between epipolar images. The method takes the independent physical properties of RGB information into consideration. The RGB differences from neighbouring pixels are treated independently rather than as combined values. In this method, for each pair of matching pixels, 12 RGB differences from four pairs of neighbouring pixels determine 12 prior probabilities of matching pixels in RGB colour space, while one combined value between two matching pixels is determined as a one likelihood probability in CIELab colour space. The algorithm makes three improvements, 1) treats RGB colour components independently as basic information sources; 2) reduces the direct negative impact of the corresponding points' colour differences, and increases the contribution of neighbouring pixels' colour differences; and 3) uses a fixed window working in the first-order Markov Random Field. Improvements 1) and 2) work together to depress the interference of corresponding pixels' colour differences and improve the correct matching rate of pixels in the regional background. Improvement 3) simplifies optimal window determination.

The rest of this paper is organized as follows. Section 2 presents details of the proposed algorithm. Section 3 contains experimental data, results and evaluation. Finally, Section 4 draws conclusions with a discussion of potential future work.

## 2. THE PROPOSED ALGORITHM

For a given pair of epipolar images  $I=\{I_L, I_R\}$ , where  $I_L$  and  $I_R$  are the left (reference) and the right (target) image, the joint posterior probability of a casual pixel  $m$  in the reference image with a given disparity  $dx$  is expressed as:

$$P(m(dx) | I) = \frac{P(m(dx))P(I | m(dx))}{P(I)} \quad (1)$$

Here,  $P(m(dx))$  and  $P(I|m(dx))$  are the synthetic likelihood probability and synthetic prior probability respectively.

In the first order Markov Random Field, each matching pixel contains four neighbouring pixels; additionally, each neighbouring pixel has four neighbouring pixels. These 13 pixels build a fixed window.

---

\* Corresponding author.

In this paper, the proposed algorithm takes the independent physical properties of RGB information into consideration. The RGB differences of neighbouring pixels are treated independently rather than as combined values. 12 RGB differences from four pairs of neighbouring pixels determine 12 prior probabilities in RGB colour space, while one combined value between two matching pixels is determined as a likelihood probability in CIELab colour space.

Thus, (1) can be rewritten as:

$$P(m(dx) | I) \propto P_{CIELab}(I | (m(dx))) \prod_{s \in \{r, g, b\}} P_s(m(dx)) \quad (2)$$

A set of posterior probabilities containing all the possible values for reference pixel  $m$  were determined. Then, the largest posterior is the optimal result.

## 2.1 Likelihood Probability

According to (2), the synthetic likelihood probability  $P(m(dx))$  is:

$$P(m(dx)) \propto P_{CIELab}(I | (m(dx))) \quad (3)$$

For matching pixels  $m$  and  $m(dx)$ , the colour difference between them in CIELab colour space can be combined as:

$$\Delta m(dx) = \sqrt{(L^*(m) - L^*(m(dx)))^2 + (a^*(m) - a^*(m(dx)))^2 + (b^*(m) - b^*(m(dx)))^2} \quad (4)$$

Then,  $\Delta m(dx)$  is convert to a proper value as:

$$P(m(dx)) = \exp(-D_1^{-1}(\Delta m(dx))) \quad (5)$$

(5) indicates that  $P(m(dx))$  is determined by parameter  $D_1$  and the combined value of the matching pixels' colour differences in CIELab colour space. For the same colour differences between matching pixels, enlarging  $D_1$  reduces the sensitivity of the corresponding pixels' actual differences.

## 2.2 Prior Probability

In this paper, we solve prior probabilities in RGB colour space. There are three steps:

### 2.2.1 The difference initialization of neighbouring pixels.

For a casual matching pixel,  $m$  and  $m(dx)$ , assuming that a casual pixel  $n$  exist in their fixed window. The differences between  $n$  and  $n(dx)$  in RGB colour space are expressed as:

$$\Delta_s(n(dx)) = |I_{L(s)}(n) - I_{R(s)}(n(dx))| \quad s \in \{r, g, b\} \quad (6)$$

Here, a simple unitary processing is used to converted colour differences to values in interval [0, 1]. The values of RGB colour differences are now defined as

$$P_{r/g/b}(n(dx)) \propto \exp(-(\Delta_{r/g/b}(n(dx)) / D_2)) \quad (7)$$

The colour differences of neighbouring pixels work in RGB colour space, unlike the per-pixel scale-invariant feature transform (per-pixel SIFT, Liu, 2011) or other color-invariant space algorithms. The per-pixel SIFT assigns a 128-dimensional vector to each pixel as its basic matching feature, where too many dimensions makes it difficult to set parameters. The colour space used in other colour-invariant space methods is also composed by three vectors; however, those three components do not have same range or weigh, and thus they cannot be converted to a proper value using a single parameter. In contrast, the RGB colour space is simple and avoids the difficulties when many parameters must be set to convert each component to its proper value in the other colour-invariant space methods.

### 2.2.2 Computing Priors.

In the proposed algorithm, prior probabilities are determined by 12 RGB differences of four connected neighbouring pixels. Assuming the prior sent from the neighbouring pixel  $q$  to central pixel  $m$  in colour  $s$  with a given value  $dx$  is  $P(m, q, s, dx)$ , then, the synthetic prior probability of joint posterior probability is expressed as:

$$P(I | m(dx)) = \prod_{s \in \{r, g, b\}} P_s(I | (m(dx))) = \prod_{s \in \{r, g, b\}, q \in N(m)} P(m, q, s, dx) \quad (8)$$

Here, the prior  $P(m, q, s, dx)$  is solved as:

$$P_{s \in \{r, g, b\}, q \in N(m)}(m, q, s, dx) = \exp(-D_2^{-1} \Delta_s(q(dx))) \prod_{p \in N(q) \setminus m} P(q, p, s, dx) \quad (9)$$

### 2.2.3 Priors normalization.

Considering the discontinuous boundary, a parameter  $\lambda$  is set to compel the discontinuous area.

$$P_{s \in \{r, g, b\}, q \in N(m)}(m, q, s, dx) = \max(P(m, q, s, dx), \lambda \max_{dx \in [0, d \max]} P(m, q, s, dx)) \quad (10)$$

Here,  $d \max$  is the given searching range for  $m$ .

The parameter  $\lambda$  have two choices:

$$\lambda = \begin{cases} J_1 & (\Delta_s(m, q) > bund) \\ J_2 & (\Delta_s(m, q) < bund) \end{cases} \quad (11)$$

If a colour component distance between central pixel and its neighbouring pixel is larger than the threshold  $bund$ , a jump exists. The parameter  $bund$  must be set to a suitable value so that the inner colour jumping within a natural continuous object is ignored and boundaries between objects are identified.

$J_1$  and  $J_2$  are designed to ensure that the updated prior is within a proper range for the rest 11 priors.  $J_1$  is larger than  $J_2$  so that the algorithm enlarges the updated prior between two jumping pixels and therefore enlarges the posterior probability to compel their dissimilarity.

Then, priors are normalized as:

$$P_{s \in \{r, g, b\}, q \in N(m)}(m, q, s, dx) = (d \max + 1) \frac{P(m, q, s, dx)}{\sum_{dx \in [0, d \max]} P(m, q, s, dx)} \quad (12)$$

### 2.3 Posteriors

After 12 prior probabilities and one likelihood probability are determined, the joint posterior probability of reference pixel  $m$  with a given value  $dx$  is expressed as:

$$P_m(dx) = \exp(-D_1^{-1}(\Delta m(dx))) \prod_{s \in \{r, g, b\}, q \in N(m)} P(m, q, s, dx) \quad (13)$$

Though (13), each pixel in a reference image acquires a group of posterior probabilities for all possible values. Adjusting parameter  $D_1$  and  $D_2$  will control the different contributions of the central pixels' and their neighbouring pixels' colour differences.

For pixel  $m$ , the optimal solution is:

$$m^{MAP} = \max_{dx \in [0, d_{max}]} (P_m(dx)) \quad (14)$$

In summary, this algorithm retains the independence of RGB colour differences. This algorithm does not combine the neighbouring pixels' RGB differences, but treats all RGB differences as independent variables. This technique makes full use of RGB colour differences, with no need for additional pre-processing.

### 3. EXPERIMENTS

In this paper, the proposed fixed-window based algorithm solved two pairs of epipolar images. Figure 1 shows both two pairs of epipolar images belonging to the same flight strip collected by an unmanned aerial vehicle at a 3500\*2100 resolution with many local error sources. Many corresponding pixels in the rectified images have colour differences, especially in building roofs and on the ground.



Figure 1. Two pairs of epipolar images.

#### 3.1 The results calculated by the proposed algorithm.

In these experiments, parameters  $D_1, D_2, bund, J_1$  and  $J_2$  are all set to 80, 255, 15, 0.85 and 0.75 based on our experimental experiences. The maximum research range for reference pixels in both two epipolar image pairs is 60 pixels.

After combining all the disparities of the pixels in the reference images, two disparity space images (DSI) with the same resolution as the reference images was generated as shown in Figure 2 (pixel's gray value = reference pixel's disparity \* 4).





(b)

Figure 2: Two DSIs calculated by the proposed algorithm. (a) is the DSI between Figure 1(a) and Figure 1(b), (b) is the DSI between Figure 1(c) and Figure 1(d). Figure 1(a) and Figure 1(c) are the reference images (Pixels' gray value = reference pixel's disparity \* 4).

Obviously, both of the DSIs in Figure 2 have two main parts: ground and building roofs. For the ground, all disparities in the background are distributed evenly. For building roofs, the consistency in the inner roof shows local disparity similarity. The roof nonconformities indicate that the roofs have altitude discrepancies. All outward appearances noted are in accordance with the physical conditions in both epipolar image pairs.

### 3.2 Boundary testing.

Boundaries are common phenomena in images. The quality of the boundaries is the key to 3D reconstruction. For evaluating boundaries, two synoptic maps overlapped by reference images and their DSIs are shown in Figure 3.

Four characteristics can be seen in Figure 3:

1. The boundaries in DSI agree with the actual conditions. First, most of buildings' boundaries in DSI can match the actual structures' edges. Second, some independent platforms are isolated from background area.
2. Most of occlusion areas are correctly calculated. The disparities of occlusion pixels are in conformity with the actual continuous area although they have no corresponding pixels in the target image.
3. Some local error sources such as moved objects are isolated automatically. Some local error sources' disparities are different from the surrounding pixels and are naturally isolated after most pixels in the background are successfully matched.
4. Some matching error exists in complex areas. When dense local buildings exist, the boundaries of buildings are not easy to be identified; these lead to many incorrect matches.

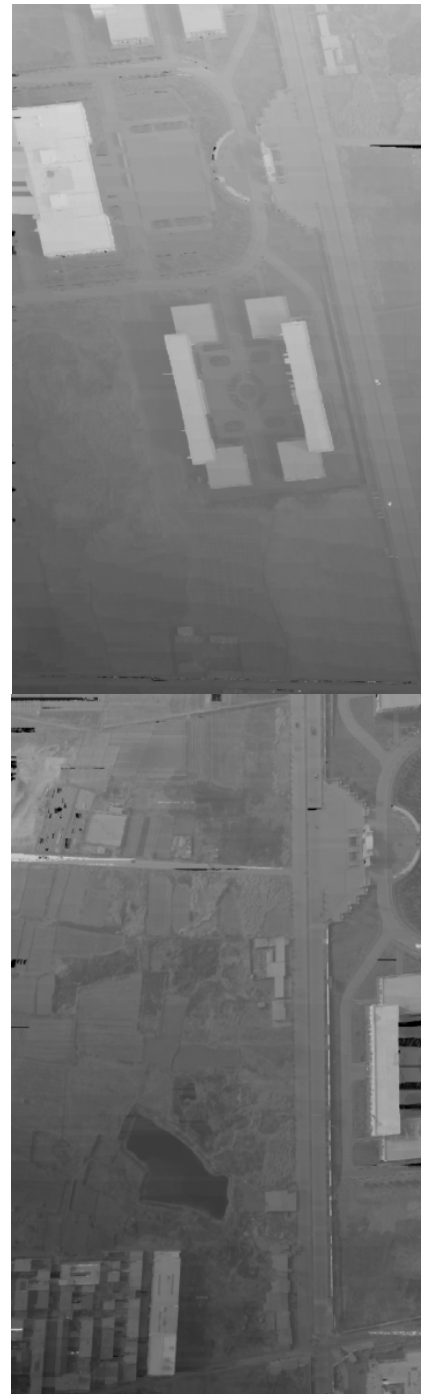


Figure 3. Two overlapped maps from the reference images and their DSIs.

### 3.3 Sampled corresponding pixels' evaluation.

In order to evaluate the accuracy of the proposed algorithm, 1,000 reference pixels were selected from Figure 1(a) and Figure 1(b) with a regular grid (20X50), then their actual corresponding pixels were manually identified from the target images. The corresponding pixels' actual results are compared with the calculated results, as shown in Table 1.

Table 1 gives error details. For image pair Figure 1(a) and Figure 1(b), 57.5% of sampled corresponding pixels are accurate, while 88.2% are controlled in 1 pixel, and 92.4% are controlled in just 2 pixels. In addition, the variance is 10.6354

since the maximum disparity error is 50 and 31 errors are more than 10 pixels. The accumulated percentages of Figure 1(c) and Figure 1(d) is 50.2%, 73.0% and 81.9%; with a variance of 12.4687.

Error (unit: pixel)	Amount	Accumulated Percentage
0	575/502	57.5%/50.2%
1	307/228	88.2%/73.0%
2	42/89	92.4%/81.9%
[3, 9]	45/95	96.9%/91.4%
[10, 60]	31/86	100%/100%
Max error	50/55 (unit: pixel)	
Variance	10.6354/12.4687 (unit: pixel)	

Table 1. The sampled corresponding pixels' statistical errors for the two epipolar image pairs from Figure 1. The left data belong to Figure 1(a) and Figure 1(b). The right data belong to Figure 1(c) and Figure 1(d).

The results from Figure 1(c) and Figure 1(d) are worse than the results from Figure 1(a) and Figure 1(b). Because of existing dense local buildings in the left lower corner of Figure 1(c) and Figure 1(d), many boundaries could not be identified. Thus, many incorrect matches cause a bigger variance and worse statistical results.

### 3.4 3D point cloud check.

Figure 4 shows 3D view of the point cloud containing 5,620,315 spatial points resulting from the two epipolar image pairs shown in Figure 1 and their DSIs. Most incorrect matches including moved objects and sensor noises were removed after the corresponding pixels' colour differences were manually limited. The spatial distribution of the 3D point cloud followed the physical conditions in two epipolar image pairs shown in Figure 1. This 3D point cloud demonstrates that the proposed algorithm is capable of generating dense corresponding pixels.



Figure 4. The 3D point cloud resulting from two epipolar image pairs in Figure 1.

## 4. CONCLUSION AND FUTURE WORK

The proposed algorithm considers independently all three colour components and works well in a simple fixed window. Experimental results show that it is capable of suppressing colour differences existing between corresponding pixels and correctly identifies most of the corresponding pixels. Thus, this algorithm can effectively solve the dense corresponding point matching problem between rectified images.

The results in Section 3 demonstrate the validity and rationality

of the proposed algorithm; however, it has some weaknesses. First, it created some wrong matches in the complex background especially in dense structure-containing areas given the difficulty in distinguishing object boundaries efficiently. Second, the pixel-leveled precision leads to pixel-leveled jumping in DSI and cause space cliff in 3D point cloud.

In the future, we plan to judge objects' boundaries by making full use of RGB's recognition of different physical properties and investigate how to improve DSI using sub-pixel estimation.

## REFERENCES

- T. Kanade and M. Okutomi, 1994. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiments. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 9, pp. 920-932.
- A. Fusiello, V. Roberto, and E. Trucco, 1997. Efficient Stereo with Multiple Windowing. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. San Juan, Puerto Rico, USA, pp. 858-863.
- Y. Boykov, O. Veksler, and R. Zabih, 1998. "A Variable Window Approach to Early Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*. vol. 20, no. 12, pp. 1283-1294.
- S.Geman and D.Geman 1984. Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.6, No.6, pp721-741.
- S.B. Kang, R. Szeliski, and C. Jinxjang, 2001. Handling Occlusions in Dense Multi-View Stereo. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Kauai, Hawaii, USA, vol. 1, pp. 103-110.
- O. Veksler, 2002. Stereo Correspondence with Compact Windows via Minimum Ratio Cycle. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1654-1660.
- O. Veksler, 2003. Fast Variable Window for Stereo Correspondence using Integral Images. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Madison, Wisconsin, USA, vol. 1, pp. 556-561.
- C. Liu, J. Yuen and A. Torralba, 2011. "SIFT flow: dense correspondence across different scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no.5, pp. 978-994.
- K.-J. Yoon and I.-S. Kweon, 2006. Adaptive Support-Weight Approach for Correspondence Search. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, pp. 650-656.

## ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant 41023001, 41021061 and 40971211.