MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Improving Accuracy of 3-D Reconstruction by Classifying Correspondences

Yusuke Takanashi† Yohei Shibata‡ Yasushi Kanazawa†

† Department of Knowledge-based Information Engineering
Toyohashi University of Technology
1–1 Hibarigaoka, Tempaku, Toyohashi, Aichi 441-8150 Japan.

‡ Denso Corporation
1–1 Showa, Kariya, Aichi 448-8661 Japan.

takanashi@img.tutkie.tut.ac.jp, kanazawa@tutkie.tut.ac.jp

Abstract

In this paper, we show that we can improve accuracies of 3-D reconstructions with uncalibrated stereo by classifying correspondences between two images. After obtaining initial correspondences by an automatic matching program, we classify the correspondences into inliers and outliers in a multi-dimensional feature space. For doing this, we introduce four quantities with respect to a corresponding pair and adopt EM algorithm and one-class SVM with a special kernel fitted to their characteristics. By real image examples, we show that we can improve the accuracies of 3-D reconstructions by classifying correspondences.

1 Introduction

3-D reconstruction using uncalibrated stereo can be applied to various fields in real world. For reconstructing 3-D structure from two images taken by such an uncalibrated stereo, we must first find corresponding pairs between the two images. We then compute the fundamental matrix from the corresponding pairs. Finally, we reconstruct the 3-D shape of the scene using the camera parameters computed by decomposing the fundamental matrix [4, 6]. However, the reconstructed shape is very sensitive to the error of the fundamental matrix. In addition, the error causes the failure in decomposing the fundamental matrix into the camera parameters even if the configuration of the cameras is not degenerate [4, 11]. Such error is mainly caused by mismatches in the correspondences. Except to make correspondences by hand, we cannot avoid including mismatches in the obtained correspondences by using any automatic matching programs [5, 7, 13]. Unfortunately, there are some mismatches that satisfy the epipolar equation in the specified degree. So, we cannot remove them completely by usual outlier detection, such as RANSAC [1] and LMedS [9].

In this paper, for improving accuracy of 3-D reconstruction with an uncalibrated stereo, we propose to distinguish the mismatches, which we call outliers, from initial correspondences in a multi-dimensional feature space using classifiers. Since there are various scenes and camera configurations, we cannot prepare any training data to distinguish them. Therefore, we introduce four features with respect to a correspondence and we apply EM algorithm and one-class SVM with a special kernel to remove outliers. We show that we can improve the accuracy of the 3-D reconstruction by classifying correspondences by real image examples.

2 One-class Support Vector Machine

The one-class support vector machine (SVM) has been proposed by Schölkopf et al.[10]. It is well known as a classification method without any training data. This classifier maps the data into a feature space and separates them from the origin with the maximum margin in the feature space.

Let $\boldsymbol{y}$ be an input data and $\boldsymbol{w}$ be the parameters of the hyperplane which separates outliers from data. We consider the following decision function

$$f(\boldsymbol{y}) = \mathrm{sign}(\boldsymbol{w}^\top \Phi(\boldsymbol{y}) - \rho). \tag{1}$$

Here, the scalar ρ indicates the distance from the origin to the hyperplane. The function $\mathrm{sign}(x)$ is the signum: if $x \geq 0$ return 1, otherwise return 0. The function $\Phi(\boldsymbol{y})$ is a non linear function that maps the data into a feature space. In order to obtain the parameters w and ρ, we must solve the following minimization:

$$\min_{\boldsymbol{w},\boldsymbol{\xi},\rho} \frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{1}{\nu N}\sum_{i=1}^{N} \xi_i - \rho, \tag{2}$$

with the constraints

$$\boldsymbol{w}^\top \Phi(\boldsymbol{y}_i) \geq \rho - \xi_i, \quad \xi_i \geq 0, \tag{3}$$

where N is the number of the input data, ξ_i, $i = 1$, ..., N are non-zero slack variables, and ν is a fraction of the outliers in the data.

By using Lagrange multiplier technique, the minimization can be rewritten by

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\alpha_i\alpha_j K(\boldsymbol{y}_i, \boldsymbol{y}_j), \tag{4}$$

subject to

$$0 \le \alpha_i \le \frac{1}{\nu N}, \quad \sum_{i=1}^{N} \alpha_i = 1. \tag{5}$$

Here, $K(\boldsymbol{y}_i, \boldsymbol{y}_j)$ is a positive semi-definite function defined by

$$K(\boldsymbol{y}_i, \boldsymbol{y}_j) = \Phi(\boldsymbol{y}_i)^\top \Phi(\boldsymbol{y}_j), \tag{6}$$

which is called a kernel. In many applications [2, 8, 10], the following radial basis function is often used:

$$K(\boldsymbol{y}_i, \boldsymbol{y}_j) = \exp\left(-\frac{\|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2}{2\sigma^2}\right), \tag{7}$$

which is called Gaussian kernel. By using the kernel (6), we can obtain

$$f(\boldsymbol{y}) = \mathrm{sign}\left(\sum_{i=1}^{N} \alpha_i K(\boldsymbol{y}_i, \boldsymbol{y}) - \rho\right), \tag{8}$$

where

$$\rho = \sum_{j=1}^{N} \alpha_j K(\boldsymbol{y}_j, \boldsymbol{y}_i). \tag{9}$$

If we know or set the fraction of the outliers ν in advance, we can compute the parameters α_i from input data by an optimization method without any training data set [10].

3 Proposed One-class SVM

In order to judge whether a correspondence is correct or not, the only constraint is the epipolar equation. However, since the epipolar constraint is not a sufficient condition but a necessary condition, some outliers are satisfy the epipolar constraint. In this paper, we try to distinguish such outliers from the correspondences in a multi-dimensional feature space. To do this, we first define four features about a correspondence and then define a special kernel fitted to their characteristics for one-class SVM.

Let P and P' be the feature points in the images I and I', respectively. We regard the pair $\{P_i, P'_i\}$ is the i-th (temporary) corresponding pair obtained by an automatic matching program. We denote these points by

$$\boldsymbol{x}_i = (x_i, y_i, 1)^\top, \quad \boldsymbol{x}'_i = (x'_i, y'_i, 1)^\top, \tag{10}$$

where (x_i, y_i) and (x'_i, y'_i) are their image coordinates of P_i and P'_i, respectively.

3.1 Feature Vector

Now, we define a feature vector $\boldsymbol{y}_i$ of the i-th corresponding pair $\{P_i, P'_i\}$ by

$$\boldsymbol{y}_i = (d_i, a_i, c_i, e_i)^\top. \tag{11}$$

We illustrate these features in detail in the following.

3.1.1 Average Local Depth: d_i

Since outliers in correspondences are often reconstructed as thorns from their neighbors [12], we should find and remove such thorns from the reconstructed 3-D shape. In this paper, we find the thorns from the projectively reconstructed shape, which need not decompose the fundamental matrix.

Let $\boldsymbol{p}_i = (p_i^1, p_i^2, p_i^3)^\top$ be the projectively reconstructed point from the pair $\{P_i, P'_i\}$. We define an average local depth of the i-th corresponding pair $\{P_i, P'_i\}$ from its neighbors by

$$d_i = \frac{1}{L} \sum_{j \in \mathcal{N}(P_i)} \Delta_{ji}, \quad \Delta_{ji} = p_j^3 - p_i^3, \tag{12}$$

where $\mathcal{N}(P_i)$ is the set of the neighbors of P_i and L is the number of the element of $\mathcal{N}(P_i)$. Here, we construct the relation of the neighborhood by by Delaunay-triangulation in the first image I.

3.1.2 Angle of the "Flow": a_i

If we superimpose the first image into the second image, we can obtain the line segments connected the corresponding pairs. We call these line segments "flows." When we rectify the images [12], the flows between the correct pairs are parallelized to the horizontal axis of the images. So, we use the angle of the flow to the horizontal axis as measure whether the correspondence is correct or not.

Let $\boldsymbol{H}$ and $\boldsymbol{H}'$ be the homographies to rectify the images I and I', respectively. The flow vector $\boldsymbol{f}_i$ is obtained by

$$\boldsymbol{f}_i = Z[\boldsymbol{H}'\boldsymbol{x}'_i] - Z[\boldsymbol{H}\boldsymbol{x}_i] = (f_i^x, f_i^y, 0)^\top. \tag{13}$$

Here, $Z[\,\cdot\,]$ is a scale normalization to make the third component 1. Then, we define the angle of the flow a_i by

$$a_i = \tan^{-1}\left(\frac{f_i^y}{f_i^x}\right). \tag{14}$$

In this paper, for rectifying images, we use the homographies obtained by the method of Sugaya et al. [12].

3.1.3 Bhattacharyya Coefficient: c_i

Correlations or residuals obtained by template matching are usually used for the first step to establish point correspondences between two images. We must not absolutely trust them, because they depend on the positions and the orientations of the two cameras. However, the pixel values of the local area around the feature point are important information for measuring a similarity of a candidate pair. So, we adopt the Bhattacharyya coefficient between two color histograms instead of that of template matching.

Let T_{P_i} and $T_{P_i'}$ be the rectangular regions centered on the point P_i and P_i', respectively. We define a similarity between the color histograms by

$$c_i = 1 - \sum_j \sqrt{H_j(T_{P_i})H_j'(T_{P_i'})}, \tag{15}$$

where $H_j(T_{P_i})$ shows the j-th hue values in the histogram constructed by the region T_p.

3.1.4 Error of the Epipolar Constraint: e_i

As well known, the epipolar equation

$$(\boldsymbol{x}_i, \boldsymbol{F}\boldsymbol{x}_i') = 0 \tag{16}$$

is satisfied if the pair P_i, P_i' is correct [3]. Here, the matrix $\boldsymbol{F}$ is rank 2 and is called the fundamental matrix. In the presence of image noise, the correct correspondences do not strictly satisfy Eq. (16). So, we need a threshold to distinguish the outliers from the correspondences. In this paper, we use the residuals of the epipolar equation as a measure of correct correspondences:

$$e_i = \frac{(\boldsymbol{x}_i, \boldsymbol{F}\boldsymbol{x}_i')}{\sqrt{\|\boldsymbol{P}_k\boldsymbol{F}\boldsymbol{x}_i'\|^2 + \|\boldsymbol{P}_k\boldsymbol{F}^\top\boldsymbol{x}_i\|^2}} \tag{17}$$

where $\boldsymbol{P}_k = \mathrm{diag}(1,1,0)$ and $\mathrm{diag}(a,b,c)$ is the diagonal matrix whose diagonal elements are a, b and c.

3.2 Proposed Kernel

The one-class SVM makes effective use of the fact that the inliers make a cluster in a feature space and outliers exists in the surrounding areas of the cluster. The Gaussian kernel (7) is often used as the kernel of one-class SVM. However, in our case, we do not adopt the Gaussian kernel from the two reasons: (1) the features d_i, a_i, and e_i can be regarded as Gaussian (Fig. 1 (a)), but the feature c_i is not Gaussian (Fig. 1 (b)); (2) the scale of the features are very different in not features but scenes. So we need normalize each feature by some scalings. Therefore, we propose a new kernel

$$K(y_i, y) = \exp\left(-\|\boldsymbol{V}^{-\frac{1}{2}}(\boldsymbol{P}\boldsymbol{y}_i - \boldsymbol{y})\|^2\right), \tag{18}$$

where, the matrix $\boldsymbol{P}$ is the diagonal matrix defined by

$$\boldsymbol{P} = \mathrm{diag}(p^1, p^2, ..., p^M), \tag{19}$$

and each p^l is

$$p^l = \begin{cases} 1 \ \cdots \ \text{ if } y^l \text{ may be Gaussian,} \\ 0 \ \cdots \ \text{ if } y^l \text{ may not be Gaussian.} \end{cases} \tag{20}$$

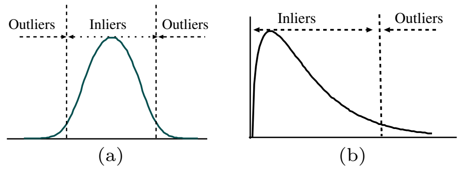

Figure 1: Types of feature distribution of $\boldsymbol{y}$.

The matrix $\boldsymbol{V}^{\frac{1}{2}}$ is

$$\boldsymbol{V}^{-\frac{1}{2}} = \mathrm{diag}\left(\frac{1}{\sigma^1}, \cdots, \frac{1}{\sigma^M}\right), \tag{21}$$

where, σ^l is the standard deviation of y^l. In this paper, for robustness, we use the semi inter-quartile range (SIQR) instead of the standard deviation.

This proposed kernel is not symmetric, so we cannot call it "kernel" strictly.

4 Removing outliers by EM algorithm

EM algorithm has been proposed as estimating unknown parameters from lack of data originally. In recently, however, it is often used for classifying data into two classes with no training data.

In this paper, we use the following two functions that map input data to feature spaces.

$$f(y_i) = \frac{1}{N}\sum_{j=1}^{N}\exp\left(-\|(\boldsymbol{V}^{-\frac{1}{2}}(\boldsymbol{P}\boldsymbol{y}_j - \boldsymbol{y}_i)\|^2\right), \tag{22}$$

$$g(y_i) = \frac{1}{N}\sum_{j=1}^{N}\left(-\|(\boldsymbol{V}^{-\frac{1}{2}}(\boldsymbol{P}\boldsymbol{y}_j - \boldsymbol{y}_i)\|^2\right). \tag{23}$$

Using these mappings, we can apply EM algorithm to distinguish outliers from correspondences.

5 Real Image Examples

Fig. 2 shows an example of a castle scene. Fig. 2 (a) shows the original images and Fig. 2 (b) shows the initial correspondences indicated by the "flow" and their 3-D reconstructions. Here, we obtain the initial correspondences by the method of Kanazawa and Kanatani [5]. In order to reconstruct a 3-D structure of the scene, we use the camera parameters decomposed from the fundamental matrix by the method of Kanatani and Matsunaga [4]. In this case, the angle between the two walls should be 90 degrees, but we see the initial 3-D reconstruction is very distorted. Fig. 2 (c) shows the correspondences obtained by the one-class SVM with the proposed kernel and their 3-D reconstructions. Fig. 2 (d) show the results obtained by the one-class SVM with the "normalized Gaussian kernel", this means we use the kernel Eq. (18) without the matrix $\boldsymbol{P}$. Fig. 2 (e) and (f) show the results obtained by the EM algorithms with Eq. (22) and Eq. (23), respectively. We see we can obtain improved 3-D

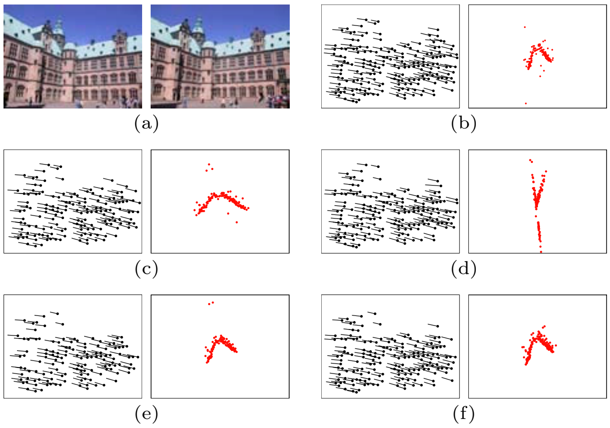

Figure 2: (a) Original images. (b) Initial correspondences and their 3-D reconstructions. (c) Results by one-class SVM with the proposed kernel ($\nu = 0.15$). (d) Results by one-class SVM the normalized Gaussian kernel ($\nu = 0.15$). (e) Results by EM algorithm with Eq. (22). (f) Results by EM algorithm with Eq. (23).

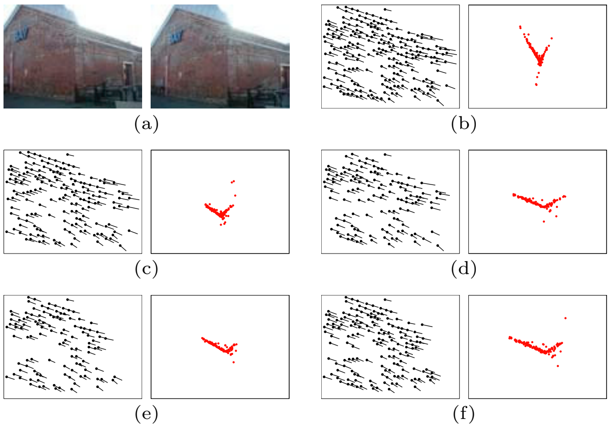

Figure 3: (a) Original images. (b) Initial correspondences and their 3-D reconstructions. (c) One-class SVM with the proposed kernel ($\nu = 0.25$). (d) One-class SVM with the normalized Gaussian kernel ($\nu = 0.25$). (e) EM algorithm with Eq. (22). (f) EM algorithm with Eq. (23).

shapes by the one-class SVM with the proposed kernel and the EM algorithm with Eq. (22).

Fig. 3 and Fig. 4 show other two examples of different building scenes. As we have seen in Fig. 2, we see the one-class SVM with the proposed kernel is the best among these methods. So, we can see the proposed non-symmetric kernel is more effective than the symmetric kernel.

6 Conclusions

In this paper, we have shown that we can improve the accuracy of 3-D reconstruction with an uncalibrated stereo by removing outliers from the corresponding pairs using classifiers. We define the four quantities on each correspondence and adopt some classifiers in a multi-dimensional feature space. We have also proposed the new kernel fitted to the features of the correspondences for the classifiers.

In real image examples, we have shown that we can improve the 3-D reconstruction of the scene by the classifiers. We have also shown the one-class SVM with the proposed kernel is robust and stable compared with the other classification methods.

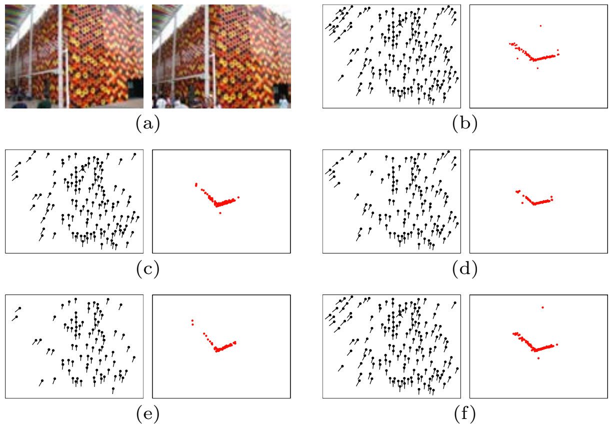

Figure 4: (a) Original images. (b) Initial correspondences and their 3-D reconstructions. (c) One-class SVM with the proposed kernel ($\nu = 0.2$). (d) One-class SVM with the normalized Gaussian kernel ($\nu = 0.2$). (e) EM algorithm with Eq. (22). (f) EM algorithm with Eq. (23).

In future works, we will explore other feature quantities about correspondences and better kernels.

References

[1] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981.

[2] I. Gondra, D. R. Heisterkamp, and J. Peng. Improving image retrieval performance by inter-query learning with one-class support vector machines. *Neural Comput. & Applic.*, 13:130–139, 2004.

[3] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University press, Cambridge, 2000.

[4] K. Kanatani and C. Matsunaga. Closed-form expression for focal lengths from the fundamental matrix. In *Proc. 4th Asian Conf. Comput. Vision*, pages 128–133, Taipei, Taiwan, January 2000.

[5] Y. Kanazawa and K. Kanatani. Robust image matching preserving global consistency. In *Proc. 6th Asian Conf. Comput. Vision*, pages 1128–1133, Jeju Island, Korea, January 2004.

[6] M. I. A. Lourakis and R. Deriche. Camera self-calibration using the kruppa equations and the svd of the fundamental matrix: The case of varying intrinsic parameters. In *INRIA*, 2000.

[7] D.G. Lowe. Distinctive image features from scale-invariant keypoint. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[8] M. J. Quinlan, S. K. Chalup, and R. H. Middleton. Application of svms for colour classification and collision detection with aibo robots. In Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[9] P. J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Wiley, New York, 1987.

[10] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13:1433–1471, 2001.

[11] P. Sturm, Z. L. Cheng, P. C. Y Chen, and A. N. Poo. Focal length calibration from two views: method and analysis of singular cases. *Comput. Vis. Image Understand.*, 99:58–95, 2005.

[12] Y. Sugaya, K. Kanatani, and Y. Kanazawa. Generating dense point matches over two images using epipolar geometry. *Memoirs of the Faculty of Engineering, Okayama University*, 40:44–57, January 2006.

[13] Z. Zhang, R. Deriche, O. Faugeras, and Q. -T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell.*, 78:87–119, 1995.