Automatic Feature Matching Using Coplanar Projective Invariants For Object Recognition

Wen-Jing Li, Tong Lee, and Hung-Tat Tsui
Dept. of Electronic Engineering, The Chinese University of Hong Kong
{wjli, tlee, httsui}@ee.cuhk.edu.hk

Abstract

The correspondence problem has been difficult for object recognition. Typically, good correspondence can only be obtained if the epiploar geometry is known as in the case of stereo image pair, or if the input images are in a video sequence. This paper proposes a new algorithm to incorporate projective invariants for solving the correspondence problem of images taken by an uncalibrated camera. As a result, the displacement of the object in the images is not required to be small. The proposed algorithm uses the cross ratio of coplanar points, an invariant from projective geometry to form the compatibility constraint, which is embedded into an energy function. Although the energy function is in high order form, we show that it can be solved using a standard second order Hopfield network, by taking advantage of the neighborhood information in the data. The proposed method has been tested on a series of real images and performs well.

1. Introduction

Reliable feature correspondences can be obtained if camera geometry or the epipolar constraint is known or for very similar input images where the correspondence problem becomes one of tracking. However, a system that does not require known camera positions and could find matches from just a few images from different perspective transformations would allow great flexibility and broad application, particularly in object recognition field.

Cross ratio is the most essential invariant with respect to projective transformations, and its perspective invariance has attracted attention of many researchers with a view to applications to object recognition from perspective images. In [4], the index function is used to select models from a model base and is constructed from projective invariants based on algebraic curves and a canonical projective coordinate frame. Successful recognition is demonstrated despite partial occlusion of the objects. Some invariant representations [5] could also be derived from the cross ratios to speed up the matching process. Most of these methods are based on the indexing method, which the matching is performed by an index function and searching in some predefined table.

Since Hopfield and Tank proposed the Hopfield network for the traveling salesman problem [8], many engineering problems have been formulated as optimization problem in which an energy function is minimized. The customary approach is to formulate the original problem as one of energy minimization and then to use a proper relaxation network to find minimum of this function. Such solutions are attractive because they offer the advantage of parallel analog VLSI implementations. Many vision problems have also been solved in this way, such as stereo image matching [2], motion estimation [1], and so on. In [2], a Hopfield network was employed for solving the global stereo matching problem using edge segments. A five-order relaxation network is proposed in [1] to find the feature correspondences for motion estimation by taking advantage of some good initial guess. The above matching processes are based on similar input images, such that the motion between the images is small.

In this paper, we propose a new algorithm for solving the well-known correspondence problem for uncalibrated camera. The object displacement is not necessary to be small. The objects are first represented as a set of feature points, such as the dominant points extracted from the outside contour of the object, or the convex hull of scattered feature points. A cost function to establish the correspondences between the feature points of a model image with a scene image, considering both the feature correspondence consistency and the projective constraint between them is then derived. The projective constraint is formulated with the well known projective invariance---the cross ratio of five coplanar points. Therefore, the correspondence problem under projective transformation has been cast as an inexact graph matching problem and formulated in terms of constraint satisfaction, which can be mapped onto a network where the nodes are the hypotheses and the links are the constraints. The network is then employed to select the optimal subset of hypotheses, which satisfies the given constraints. A second order Hopfield network is employed in this paper, such that the convergence of the network can be guaranteed. This is made possible by utilizing the neighborhood information in the data. Based on the correspondence detected, the projective transformation, which can be represented as a homography matrix, can be

recovered to map the model object into the scene object domain, or vice versa.

The remaining of the paper is organized as follows. The preliminary background covering the projective invariance and convex hull are introduced in Section 2. Section 3 shows how the projective invariant matching can be mapped to a modified Hopfield network to find the correspondences. Section 4 is devoted to recover the projective transformation from the correspondence points detected by the network. The experimental examples with real images are given in Section 5. Finally, conclusions are presented in Section 6.

2. Projective Invariance and Convex Hull

2.1. Invariance on five coplanar points

Given five points on a plane, $\boldsymbol{p_i}=(x_i, y_i, z_i)$, $i=1, \ldots, 5$, a square matrix is defined by three of the five points:

$$m_{ijk} = \begin{pmatrix} x_i & x_j & x_k \\ y_i & y_j & y_k \\ z_i & z_j & z_k \end{pmatrix} \tag{1}$$

It can be shown that there are two functionally independent ratios [6] of the determinants of the matrices, m_{ijk}, of the five points which are invariant for projective transformations and the homogenous scalar factor:

$$I_1 = \frac{\det(m_{431})\det(m_{521})}{\det(m_{421})\det(m_{531})} \tag{2}$$

$$I_2 = \frac{\det(m_{421})\det(m_{532})}{\det(m_{432})\det(m_{521})} \tag{3}$$

Note that the three points in a triple cannot be collinear, otherwise the determinant of the point matrix m_{ijk} becomes singular and the corresponding invariant is undefined. It can also be proved that the variances of the five coplanar point invariants are proportional to their magnitude. This implies that given estimates of the value of the invariant and of the error introduced by the sensor and feature extraction scheme, we can estimate the accuracy of the invariant. Therefore, this property of projective invariance can be used as the compatibility constraints embedded the energy function to check the global consistency in feature matching.

2.2. Convex hull

In this section, we introduce the concept of convex hull and its properties. For a set of points in the plane, the convex hull is the smallest convex object containing all the points. The convex hull bounds the set of points from the outside. It possesses very attractive properties that make it suitable for shape representation and analysis. 1) It has uniqueness. 2) It has computational efficiency, the upper bound of the computational complexity associated with finding the convex hull of N points is of order *O(NlogN)* [7]. 3) It has projective invariance, which means that the convex hull of a data set under projective transformation is simply the projective transformed convex hull of the data before the transformation [5]. 4) It has local controllability, which means that when feature points are either added to or subtracted from the original data set, the convex hull is only locally affected, which is useful for handling occluded object recognition. 5) Moreover, the ordering of the vertices of the convex hull is readily available. This property is very useful to introduce the compatibility constraints among five pairs of points in the second order Hopfield network.

3. Hopfield Network for Projective Invariant Matching

Hopfield network has been used in solving many optimization problems [8]. Now we design the network structure for projective invariant matching. First, a model graph is constructed by extracting dominant feature points of the model image as the nodes of the graph, and a second graph is constructed from the scene image taken from different and unknown viewing angle of the same object, which is called the scene graph. For graph matching, Hopfield network can be considered as a 2D array. If the model graph has *M* nodes and the scene graph has *S* nodes, the number of neurons in the network will be *M***S*. The final state of each neuron represents whether the corresponding node in the model graph matches the node in the scene graph or not. The network configuration can be seen in Fig.1. The objective of the network is to optimize the defined energy function until it reaches a minimum as the neurons converge to stable states.

We define the our energy function for projective invariant matching to be

$$\begin{aligned} E = {} & \frac{A}{2}\sum_i\sum_l\sum_{m\neq l} v_{il}v_{im} + \frac{B}{2}\sum_i\sum_l\sum_{j\neq i} v_{il}v_{jl} \\ & - \frac{R}{2}\sum_i\sum_l v_{il} - \frac{D}{2}\nu\sum_i\sum_l {}_{il}v_{il} \\ & - \frac{D}{2}(1-\nu)\sum_i\sum_l\sum_{j\neq i}\sum_{m\neq l} c_{iljm}v_{il}v_{jm} \end{aligned} \tag{4}$$

where A, B, R and D are constants. v_{il} is the output state of neuron (i,l). If the i^{th} node from the reference image matches the l^{th} node from the test image, v_{il} will be 1; otherwise, it will be 0. ϖ, where $0\leq\varpi\leq1$, is the control parameter.

The first two terms of Equation (4) are uniqueness constraints, which force that at most one neuron can be active in each column and row of the network. The third constraint has to be included to avoid the system being trapped to the degenerated state in which all neurons are

inactive. The last two terms are the compatibility constraints that are used to measure the strength of the compatibility between the nodes from the model image and the scene image. The fourth term only considers the information of unary properties of the feature points detected and the last term uses the information of relational properties between the model graph and the scene graph, and in this case, the projective invariant constraints are considered.

The unary constraint is defined as:

$$c_{il} = 2/(1+e^{(e-q)/l}) - 1 \quad (5)$$

where $e = | fm_i - f_l |$. λ is the temperature constraint, determining the steepness of the function. θ is a threshold for the system to tolerate addictive noise. fm_i and f_l are unary properties of the feature points. They can be selected as the convexity and concavity of the points from the outside contour, or the radiometric similarity of the points. If no such kind of unary properties is available, (e.g. for the convex hull, all the points are convex) we just keep it unchanged, because the relational constraint plays the key role in the matching.

The relational constraint is defined as:

$$c_{iljm} = 2/(1+e^{(e-q)/l}) - 1 \quad (6)$$

where

$$e = \frac{1}{4}\left(|I_{m1} - I_{s1}| + |I_{m2} - I_{s2}| + \left|\frac{1}{I_{m1}} - \frac{1}{I_{s1}}\right| + \left|\frac{1}{I_{m2}} - \frac{1}{I_{s2}}\right| \right) \quad (7)$$

and

$$I_{m1} = \frac{\det(m_{pki})\det(m_{xji})}{\det(m_{pjx})\det(m_{xki})}, I_{s1} = \frac{\det(m_{qnl})\det(m_{yml})}{\det(m_{qml})\det(m_{ynl})}$$

$$I_{m2} = \frac{\det(m_{pji})\det(m_{xkj})}{\det(m_{pkj})\det(m_{xji})}, I_{s2} = \frac{\det(m_{qml})\det(m_{ynm})}{\det(m_{qnm})\det(m_{yml})}$$

According to projective invariance described in Section 2, for projective invariant matching, at least five pairs of nodes are needed to compute the relational constraint. Therefore, we use the relational properties between quintuple set of nodes *(i, j, k, x, p)* of the model graph and *(l, m, n, y, q)* of the scene graph as the compatibility constraint. e is the average difference of projective invariants over five points between the model graph and the scene graph. It can be proved that this constraint is symmetric (c_{iljm}=c_{jmil}) for the variable indices *i, j* and *l, m*, which is the necessary condition for the network to converge. This relational constraint means that if the value of e is smaller than the threshold q, there exists a projective transformation mapping the set of nodes from the model to the scene image, c_{iljm} approaches +1, otherwise such transformation does not exist, and c_{iljm} approaches -1.

Now the next question is how to select the reference nodes *k, x, p* in the model graph and *n, y, q* in the scene graph, such that when the i^{th} node and j^{th} node in the model match the l^{th} node and m^{th} node in the scene respectively, the k^{th}, *xth*, and p^{th} node most probably match the n^{th}, *yth* and q^{th} node respectively. The feature points used to form the graph in the network can be extracted either along the shape of the object or its convex hull in the image, thus they can be arranged in order. Therefore, we can always select the adjacent nodes of *i* and *j* as the nodes *k, x, p* in the model graph. Similarly, select the adjacent nodes of *l* and *m* as the nodes *n, y, q* in the scene graph.

From Equation (4), it can be seen that when $v=1$, the last term of the energy function is zero, the network only uses the information of unary features. When v is gradually reduced, the weight of last term becomes larger and larger. When v is reduced to zero, the network only uses the information of relational properties. We can adopt this approach to integrate the local and relational properties in the Hopfield network.

Equation (4) can be rewritten as the Liapunov function form of a Hopfield network [8] to obtain the updating equation, more details can be found in [13]. Therefore, the matching algorithm of second order network for projective invariant matching can be described as follows:

Step 1: Set the initial state of the network, and the control parameter v is set to be 1.

Step 2: Update the state of the network till a stable output state is achieved.

Step 3: Reduce the control parameter v with a small value: $v = v$ - *step*, check if $v>0$, if yes, go to Step 2; otherwise, go to Step 4.

Step 4: Output the matching results.

4. Finding the Projective Transformation between Correspondence Points

After finding the correspondence points between two views by the proposed Hopfield network, we can compute the projective transformation between them. If the points that are put into correspondences are produced by the visual features situated in a plane, there exists an analytic transformation between the two projective planes [11] completely specified by a 3*3 transformation matrix **H**,

$$\lambda_i \mathbf{P}_i^s = \mathbf{H}\mathbf{P}_i^m \quad (8)$$

where **H** is named homography (or collineation). **H** is only defined up to a scale factor, which means that one element of **H** may be set to unity, $\mathbf{H}_{3,3} = 1$. Therefore, a minimum of four pairs of correspondence points is required to solve the 8 free components of **H**.

Since we need a hypothesis verification scheme to verify the detected correspondences, and meanwhile, delete the spurious matches generated by the network, a post clustering algorithm similar to [10] was employed to estimate **H**. The algorithm finds the local maxima by voting in the parameter space.

5. Experimental Results

5.1. Projective invariant shape recognition

First, the proposed method has been evaluated with shape images taken from different and unknown viewing positions. The arrow symbol images shown in Fig.2 were taken in our laboratory with a digital camera. The images were segmented by intensity thresholding. The feature points were chosen as extreme curvature points along the outside contour of the objects, and they were extracted and labeled in clockwise manner, by applying a similar algorithm proposed in [9]. The matching results are summarized in Table 1.

Table 1. Matching results of arrow images in Fig.2

Model	Scene	Energy	Mat.	Err.
Fig.2a	Fig.2b	-36.41	9	0
Fig.2a	Fig.2c	-34.59	9	0
Fig.2a	Fig.2d	-15.99	6	0

In the table, the third column is the final energy value when the network converges. The fourth column is the number of correspondences found by the matching process, among which, the last column denotes the number of false matches. From the table, all the experiments can find the correct correspondences between the model and the scene without any wrong match, even when the scene graphs are occluded, such as Fig.2 (d).

In this set of experiments, the network converges within 3 seconds on a Sunsparc 10 (for the un-occluded cases, it is much faster). The numbers of model nodes and scene nodes are both 9. The transformed models are overlaid onto the scene in Fig.3, according to the estimated **H**. The dashed lines in Fig.3 denote the scene contours, and the solid lines denote the transformed contours of the model object. It can be seen that the transformed model contours almost perfectly match the scene contours.

5.2. Projective invariant matching by convex hull

Next, we evaluate the proposed method by matching discrete point set using convex hull. Fig.4 is a set of real images taken in our department with large differences of viewing positions. Harris corner detector [12] was used to extract features from these images. The quickhull [7] algorithm was employed to find the convex hull for each image. The matching details are listed in Table 2.

Table 2. Matching results of "EE society" images in Fig.4

Model /M	Scene/S	Ene.	Mat.	Err.	Tmat.
Fig.4a/12	Fig.4b/11	-27	8	0	120
Fig.4a/12	Fig.4c/12	-26	10	0	129
Fig.4a/12	Fig.4d/12	-30	10	1	73

In the table, M and S denote the number of nodes in the model and the scene respectively. "Tmat." means the number of correct matches detected for the full set of corner points based on the homography matrix **H** estimated from the convex hull. The error is no more than 2 pixels. From the table, there is one wrong match in the third experiment. However, they can be eliminated with the subsequent post-clustering algorithm. The transformed models including the convex hull and other corner points are overlaid onto the scenes, which are shown in Fig.5. The "+" marks denote the corner points in the scene image, and the dashed lines denote the convex hull computed for the scene images. While the "o" marks denote the corner points in the transformed models, and the solid lines denote the convex hull computed for the transformed models.

From Fig.5, it can be seen that by using the convex hull of a set of discrete points to perform matching process, we still can get a good approximation of the projective transformation. Of course, if more accurate results are required, we can use the points inside the convex hull to refine the projective transformation further. This set of experiments usually takes no more than 15 seconds to converge on a Sunsparc 10.

5.3. Experiments on 3D object recognition

In this set of experiments, the pictures shown in Fig.6 are images of a "Vita" drink pack viewed from different and unknown positions. Unlike the previous two sets of experiments where the objects are either flat or resided on a 2D plane in a 3D space, the feature points on the "Vita" drink are non-planar. This set of experiments is considered to examine the 3D structure effects on the 2D projective invariant matching.

For the three scene images in Fig.6 (b), (c) and (d), the matching process can find all the six correct matches along the 3D shape of the drink box immediately without any false match. The transformed models by the estimated **H** are overlaid on to the scenes, which are shown in Fig.7. The experimental results show that the proposed matching

algorithm also works for the non-planar object if the perspective effect is not so strong.

6. Conclusions

This paper proposes a neural network solution for automatic feature matching using coplanar projective invariants and convex hull. The problem is formulated as a minimization process, in which the energy function includes the constraints based on projective invariants of five coplanar points. A modified Hopfield network has been adopted to integrate both the unary properties and the relational properties of the feature points. By taking advantage of the neighborhood information in the data (shape or convex hull), this energy function can be solved by a second order Hopfield network such that the convergence can be guaranteed.

The experiment results show that the proposed method can handle the correspondence problems of planar objects, or non-planar objects if the perspective effects are not so strong. Prior information about the epipolar geometry is not required in the formulation. We have not assumed that the images have been calibrated, nor the objects have only been moved slightly between the images. Therefore, the proposed method has great potentials in various applications, such as robot navigation, and object recognition.

Acknowledgement

The project is partially supported by the CUHK Direct Grant 01/02.

References

[1]. A. Branca, E. Stella and A. Distante, "Feature matching constrained by cross ratio invariance", Pattern Recognition, Vol. 33, pp. 465-481, 2000.

[2]. G. Pajares, J. M. Cruz and J. Aranda, "Relaxation by Hopfield network in stereo image matching", Pattern Recognition, Vol.31, No. 5, pp. 561-574, 1998.

[3]. V. Tsonis, K. V. Chandrinos and P. Trahanias, "Landmark_based navigation using projective invariants", in Proc. of IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems, pp.342-347, 1998.

[4]. C.A. Rothwell and A. Zisserman, "Planar object recognition using projective shape representation", Inter. Journal of Computer Vision, Vol. 16, pp. 57-99, 1995.

[5]. Peter Meer, S. Ramakrishna and R. Lenz, "Correspondence of coplanar features through P^2-invariant representations", in Proc. of Inter. Conf. on Pattern Recognition, pp.196-199, 1994.

[6]. J. L. Mundy and A. Zisserman, Geometric Invariance in Computer Vision, MIT Press, 1992.

[7]. C. B. Barber, D. P. Dobkin and H Huhdanpaa, "The quickhull algorithm for convex hulls", ACM Trans. on Mathematical Software, Vol.22, No.4, pp469-483, 1996.

[8]. J. J. Hopfield and D. W. Tank, "Neural computations and decisions in optimization problems", Biological Cybernetics, Vol. 52, pp.141-152, 1985.

[9]. N. Ansari and E.J. Delp, "On detecting dominant points", Pattern Recognition, Vol.24, No.5, pp.441-451, 1991.

[10]. P.N. Suganthan, E. K. Teoh and D.P. Mital, "Pattern recognition by graph matching using the potts MFT neural networks", Pattern Recognition, Vol.28, No.7, pp.997-1009, 1995.

[11]. D. Sinclair, "Quantitative planar region detection", Inter. Journal of Computer Vision, Vol.18, No. 1, pp.77-91, 1996.

[12]. C. Harris, "A combined corner and edge detector", in Proc. of 4th Alvey Vision Conf., pp.147-154, 1988.

[13]. Wen-Jing Li and Tong Lee, "Object recognition by sub-scene graph matching", in Proc. of Intern. Conf. On Robotics and Automation, pp.1459-1464, 2000.

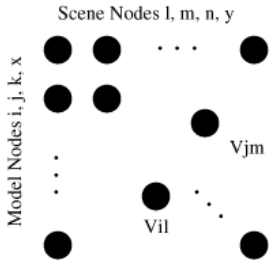


Fig.1. The Hopfield network used to generate graph isomorphism

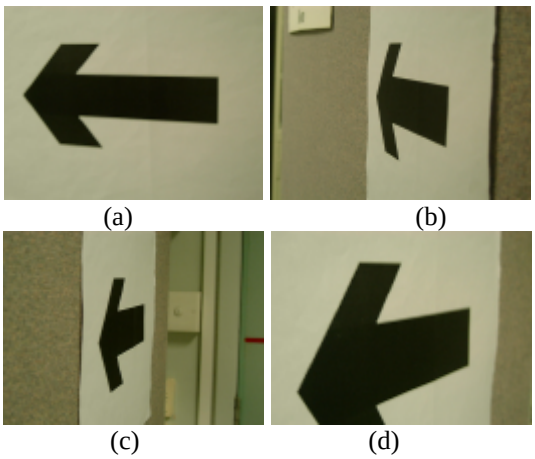
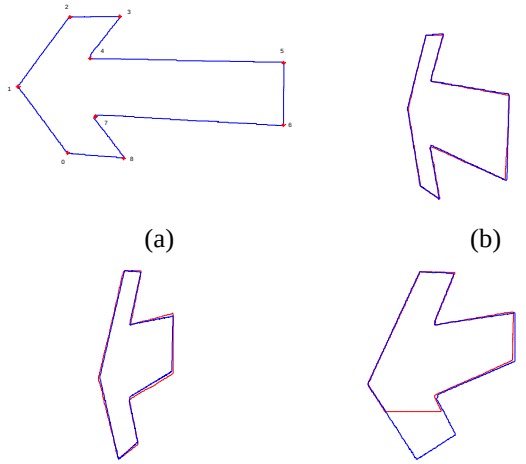


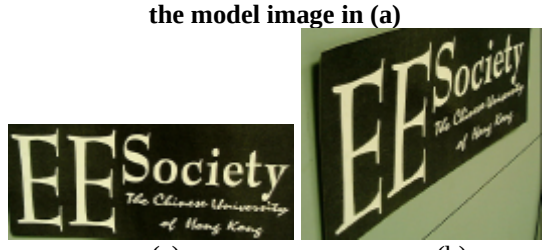
Fig.2. 2D symbol images, (a) is the model image



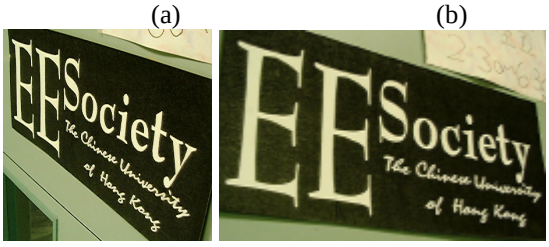
(a) (b)

(c) (d)

Fig.3. Matching results of 2D symbol images against the model image in (a)

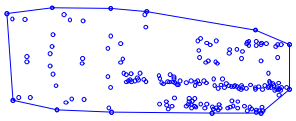


(a) (b)

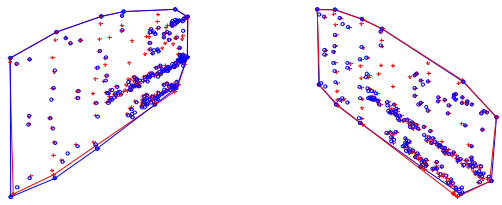


(c) (d)

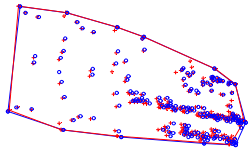
Fig.4. "EE society" label images, (a) is the model



(a)

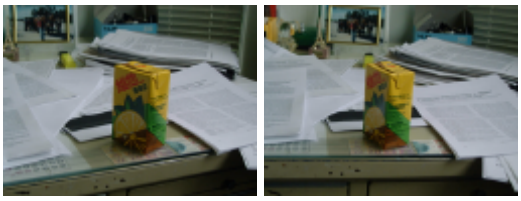


(b) (c)

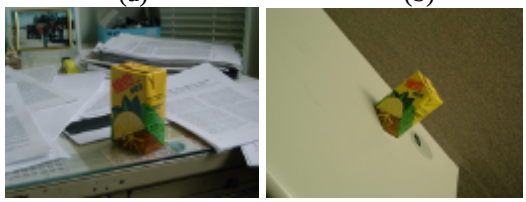


(d)

Fig.5. Matching results of "EE society" images against the model image in (a)

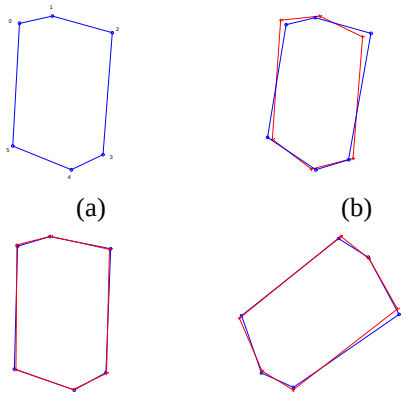


(a) (b)



(c) (d)

Fig.6. "Vita" drink images, (a) is the model images



(a) (b)

(c) (d)

Fig. 7. Matching results of "Vita" drink images against the model image in (a)