

Correspondence Propagation with Weak Priors

Huan Wang, Shuicheng Yan, *Member, IEEE*, Jianzhuang Liu, *Senior Member, IEEE*, Xiaoou Tang, *Fellow, IEEE*, and Thomas S. Huang

Abstract—For the problem of image registration, the top few reliable correspondences are often relatively easy to obtain, while the overall matching accuracy may fall drastically as the desired correspondence number increases. In this paper, we present an efficient feature matching algorithm to employ sparse reliable correspondence priors for piloting the feature matching process. First, the feature geometric relationship within individual image is encoded as a spatial graph, and the pairwise feature similarity is expressed as a bipartite similarity graph between two feature sets; then the geometric neighborhood of the pairwise assignment is represented by a categorical product graph, along which the reliable correspondences are propagated; and finally a closed-form solution for feature matching is deduced by ensuring the feature geometric coherency as well as pairwise feature agreements. Furthermore, our algorithm is naturally applicable for incorporating manual correspondence priors for semi-supervised feature matching. Extensive experiments on both toy examples and real-world applications demonstrate the superiority of our algorithm over the state-of-the-art feature matching techniques.

Index Terms—Feature matching, weak prior, image registration, object correspondence, propagation.

I. INTRODUCTION

FEATURE matching of two objects is a fundamental problem for computer vision research, and a variety of computer vision tasks heavily rely on the feature matching results, such as object tracking [14] and recognition [18], [17], image warping [6] and stitching [10], and 3-D reconstruction [5], [19], [4]. The feature matching accuracy may be affected by various factors including feature descriptors, similarity measurements, and matching approaches.

Substantive works have been devoted to seeking the correspondences between features extracted from two images. Among them, recently Grauman *et al.* [13] considers the image

features as unordered elements in sets of different cardinalities and proposes a pyramid matching algorithm for pursuing inexact correspondences. Local feature plays an important role in this task, and the popular feature detectors, such as SIFT [17], salient region detector [15], as well as scale and affine invariant interest point detector [20], tend to output interest points or regions in a structured way. Also, it is observed that the salient points and SIFT features extracted from the images with similar structures often share similar local spatial distributions. Thus, the feature location also conveys important information for feature matching. The works in [9], [23] and [21] present approaches for utilizing structure information. They formulate the feature matching problem with integer quadratic programming (IQP) or Semidefinite Programming (SDP) techniques and, hence, severely suffer from the high computational cost. Leordeanu *et al.* [16] proposes a spectral analysis method for promoting feature matching accuracy with the geometric structure information and designs an iterative procedure to eliminate the conflicts among the derived correspondences. [12] adds affine constraints to the spectral matching formulation and proposes a normalization procedure to improve the matching accuracy. Recently, [24], [25] use “window-like” neighborhood searching and design propagation algorithms for stereo image matching.

One common issue encountered by all above feature matching algorithms is that the top few matches with the highest similarities are often very accurate, but the matching accuracy falls rapidly when the desired match number increases, especially for data with noises. A demonstration of the deterioration of matching quality as the number of matches increases is shown in Fig. 1. Another issue arising in real-world applications is that the unsupervised feature matching algorithms often cannot provide sufficiently accurate results for the subsequent applications such as image stitching and object recognition. A natural question is how to incorporate extra clues for promoting feature matching performance. In this work, we present a solution for feature matching with the *reliable correspondence priors*, from the top few reliable correspondences obtained by either conventional feature matching algorithms or manual labeling.

First, the relative geometric relation of the feature pairs within an image is encoded as a spatial graph, and the matching assignments are considered as the vertices of the product graph constructed from two spatial graphs of the images to be matched. Then, based on the these spatial relations, the assignment neighborhoods are defined on the product graph and the point-to-point matchings are then propagated from those reliable correspondences to the remaining ones. Finally, we deduce an efficient closed-form solution for the feature matching problem by

Manuscript received September 12, 2007; revised June 26, 2008. Current version published December 12, 2008. This work was supported in part by a grant from the Research Grants Council of the Hong Kong SAR, China (Project No. CUHK 414306), and in part by AcRF Tier 1 Grant of R-263-000-464-112, Singapore. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Elad.

H. Wang is with the Department of Computer Science, Yale University, New Haven, CT 06511 USA (e-mail: huan.wang@yale.edu).

S. Yan is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore (e-mail: eleyans@nus.edu.sg).

J. Liu is with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: jzliu@ie.cuhk.edu).

X. Tang is with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, and also with Microsoft Research Asia, Beijing, China, 100080 (e-mail: xtang@ie.cuhk.edu.hk; xitang@microsoft.com).

T. S. Huang is with the Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61820 USA (e-mail: huang@ifp.uiuc.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2008.2006602

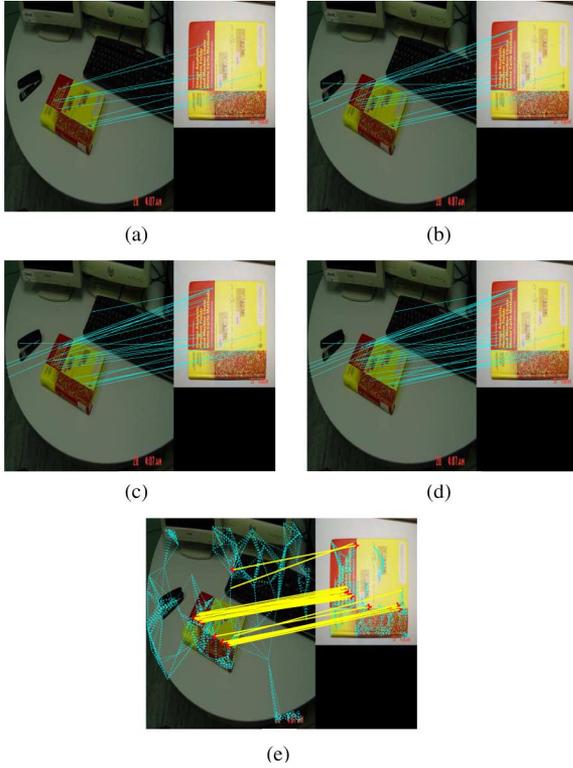


Fig. 1. Deterioration of matching quality as the number of matches increases. (a)–(d) Matching results of the traditional algorithm proposed in [17]. We vary the parameters so that the number of matches from (a) to (d) increases. (e) iMatching performance of the automatic RCP algorithm. Note the number of matches in (e) and (d) are the same but the matches derived by our algorithm are more geometrically consistent and the number of mismatches is smaller.

ensuring both spatial consistency and feature similarity agreements.

The works in [9], [16], and [21] try to employ the feature location information for matching, while our work in this paper differs from them in that we make full use of the information provided by those reliable correspondences. Moreover, benefiting from the propagation property, our framework is easy to incorporate human interactions for the guidance of correspondence searching. An illustration of the whole framework for correspondence propagation from reliable correspondence priors is displayed in Fig. 2.

Here, we would like to highlight some aspects of our proposed Reliable Correspondence Propagation (RCP) algorithm.

- 1) RCP makes full use of the prior information of the sparse reliable correspondences, and is naturally applicable for incorporating the interactive manual labeling to further promote feature matching accuracy in a semi-supervised way.
- 2) The algorithmic objective provides a unified formulation that employs both the categorical product graph constructed from two spatial graphs for characterizing spatial coherency and the bipartite similarity graph for representing feature similarity agreements.
- 3) A closed-form solution is deduced with comparably low computational cost, and, hence, our algorithm is applicable to real-world image registration problems.

II. PROBLEM FORMULATION AND SOLUTION

A. Notations and Graph Construction

The two sets of features, e.g., extracted from SIFT [18], within two images to be matched are denoted as $\Phi^1 = \{\phi_1^1, \phi_2^1, \dots, \phi_{N_1}^1\}$ and $\Phi^2 = \{\phi_1^2, \phi_2^2, \dots, \phi_{N_2}^2\}$ with $\phi_i^k = \{f_i^k, x_i^k\}$, where f_i^k is the feature vector and x_i^k is the feature point location in the k^{th} image ($k \in \{1, 2\}$).

Let $G^k = (V^k, E^k)$ be an undirected spatial graph with vertex set V^k and edge set E^k for the k^{th} image. The vertices V^k represent the feature points extracted from the input images and the edges in E^k reflect the geometric neighboring relations among the features, and can be defined in terms of k -nearest-neighbor or an ϵ -ball distance criteria in the feature position space. We call G^k spatial graph because it essentially encodes the relative spatial positions of all the feature pairs within the input image k . In addition, an adjacency/weight matrix W^k is defined for the graph G^k . One way to compute the weight matrix is directly based on the edge information, namely

$$w_{ij}^k = \begin{cases} 1, & \text{if } x_i^k \text{ and } x_j^k \text{ are connected} \\ 0, & \text{else.} \end{cases}$$

There are also other ways for computing the similarity matrix, such as the heat kernel [7], i.e., $w_{ij}^k = e^{-\|x_i^k - x_j^k\|^2 / t}$, where $t \in \mathbb{R}$ is a parameter to define the heat kernel.

To encode the pairwise feature similarity between two sets of features, we introduce the similarity graph, denoted as a triplet $G^{12} = (\Phi^1, \Phi^2, E^{12})$. The similarity graph G^{12} is a bipartite graph, and the weight matrix S of G^{12} are computed from the cosine distances of the feature pairs measured in the feature vector space.

B. Regularization on Categorical Product Graph

The feature matching process can be considered as seeking a binary function over the product set of Φ^1 and Φ^2

$$\mathcal{M} : \Phi^1 \times \Phi^2 \rightarrow \{0, 1\}$$

where \times denotes the set product and the function value 1 means matching and 0 for mismatching. To transduce the matching assignment from the reliable correspondence priors to the other feature pairs, we first give a neighborhood definition for the matching assignments.

Definition: Suppose $\Phi^1 = \{\phi_{i_1}^1, \phi_{i_2}^1, \dots, \phi_{i_{N_1}}^1\}$ and $\Phi^2 = \{\phi_{j_1}^2, \phi_{j_2}^2, \dots, \phi_{j_{N_2}}^2\}$ are the vertices of graph G^1 and G^2 respectively. Two assignments $m_{i_1 i_2} = \{\phi_{i_1}^1, \phi_{i_2}^2\}$ and $m_{j_1 j_2} = \{\phi_{j_1}^1, \phi_{j_2}^2\}$ are neighbors iff both pairs $\{\phi_{i_1}^1, \phi_{j_1}^1\}$ and $\{\phi_{i_2}^2, \phi_{j_2}^2\}$ are neighbors in G^1 and G^2 respectively, namely

$$m_{i_1 i_2} \sim m_{j_1 j_2} \text{ iff } \phi_{i_1}^1 \sim \phi_{j_1}^1 \text{ and } \phi_{i_2}^2 \sim \phi_{j_2}^2 \quad (1)$$

where $a \sim b$ means a and b are neighbors on the corresponding graph.

Suppose binary weights are utilized. According to the definition (1), the assignment graph G^a is the categorical product graph of G^1 and G^2 , i.e., $G^a = G^1 \times G^2$, and the adjacency of the assignments can be expressed as

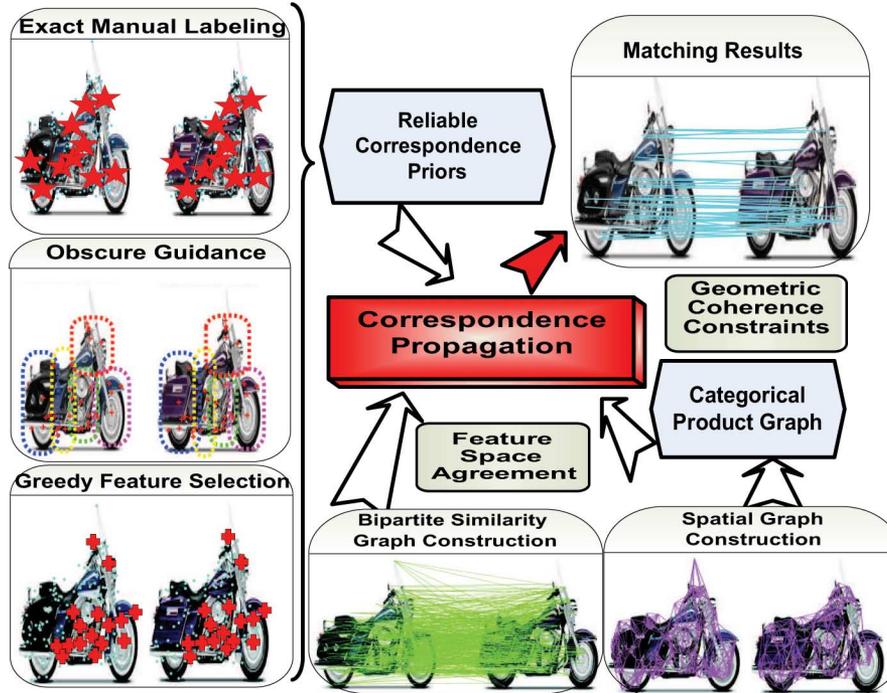


Fig. 2. Flowchart of correspondence propagation from reliable correspondence priors for feature matching.

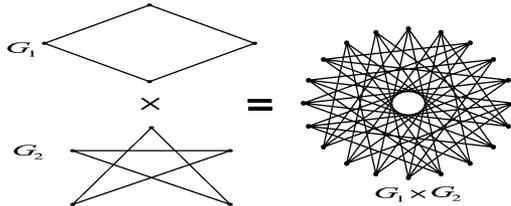


Fig. 3. Demonstration of categorical graph product: graph G_1 , graph G_2 , and their categorical graph product $G_1 \times G_2$.

$$w_{m_{i_1 i_2} m_{j_1 j_2}}^a = w_{\phi_{i_1}^1 \phi_{j_1}^1} w_{\phi_{i_2}^2 \phi_{j_2}^2} \cdot \quad (2)$$

An example of the categorical product graph is demonstrated in Fig. 3.

Defined on the space of Cartesian product set, the assignment \mathcal{M} can be regarded as a binary matrix of N_1 by N_2 , i.e.,

$$M = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1N_2} \\ m_{21} & m_{22} & \cdots & m_{2N_2} \\ \cdots & \cdots & \cdots & \cdots \\ m_{N_1 1} & m_{N_1 2} & \cdots & m_{N_1 N_2} \end{pmatrix}, m_{ij} \in \{0, 1\} \quad (3)$$

where the elements m_{ij} corresponds to the assignment of ϕ_i^1 to ϕ_j^2 . To facilitate the solution, we arrange the columns of M consequently to construct a vector \vec{M} , i.e.,

$$\vec{M} = \text{vec}(M) = [m_{11}, m_{21}, \dots, m_{N_1 1}, m_{12}, m_{22}, \dots, m_{N_1 2}, \dots, m_{1N_2}, m_{2N_2}, \dots, m_{N_1 N_2}]^T \quad (4)$$

where $\text{vec}(\cdot)$ is the vectorization operator.

Now the assignment is converted into a function on the $N_1 N_2$ dimensional vector space and, thus, the adjacency matrix W^a of the assignments is an $N_1 N_2$ by $N_1 N_2$ matrix, i.e.,

$$W^a = W^{2T} \otimes W^1 \quad (5)$$

where \otimes is the Kronecker product operator and the corresponding graph G^a is the categorical product graph of G^1 and G^2 . Note that the adjacency matrix of the categorical product graph can also be defined as $W^a = W^1 \otimes W^2$ if we rearrange the sequence of assignments while here we adopt the first definition so that the assignment arrangement is coherent with that of \vec{M} . When the W^1 and W^2 are not binary, the adjacency matrix W^a calculated from (5) can still capture the relative geometric relations of the assignments.

To introduce a spatial consistency for the assignments, we make the assumption that the neighboring vertices on the categorical product graph share similar assignment values. That is, the feature pair tends to be matched if its neighbor feature pair, i.e., the neighboring vertex in the categorical product graph, has the assignment value of 1 (stands for matching). This is quite natural in the representation of structural feature sets, since in real-world applications, the feature points that constitute certain kind of structures are often extracted together, and, thus, the features are often matched *set by set*. Emphasizing this assumption can also transduce the *reliable correspondence priors* derived from manual labeling or automatic approaches to their neighboring assignments and then the assignments are propagated along the categorical product graph until a final balance is drawn.

In our framework, the spatial consistency assumption is fulfilled by a graph Laplacian penalty item in the objective. According to the spectral graph theory [8], [22], penalizing the

Graph Laplacian leads to a solution with the locality preserving property. The Graph Laplacian can be expressed as

$$\vec{M}^T L^a \vec{M}^T = \frac{1}{2} \sum_{ij} w_{ij}^a (\vec{M}_i - \vec{M}_j)^2$$

where \vec{M}_i is the i th element of \vec{M} , $L^a = D^a - W^a$ is the Laplacian matrix of the categorical product graph and D^a is a diagonal matrix with $D_{ii}^a = \sum_j W_{ij}^a$. If \vec{M}_i and \vec{M}_j are adjacent in the graph, i.e., the w_{ij}^a is large, the minimizing of the objective will lead to a small distance between \vec{M}_i and \vec{M}_j , and then the reliable prior correspondence can be propagated along with this spatial consistency property.

C. Consistency in Feature Domain and Soft Constraints

Besides the geometric consistency, we also emphasize the coherence in the feature domain. The pairwise feature agreement is encoded by the N_1 by N_2 adjacency matrix S of the similarity graph G^{12} . The coherence of the feature similarity is then converted into the maximization of item

$$|S \odot M|_s = \vec{S} \cdot \vec{M} = \vec{S}^T \vec{M}, \text{ w.r.t. } \vec{M}_i \in \{0, 1\} \quad (6)$$

where \odot is the matrix Hadamard product, $|A|_s$ returns the sum of all the elements in matrix A , \vec{S} is the vectorization of the matrix S , and the operator \cdot is the inner product of two vectors.

Finally, for those one-to-one correspondence configurations, a soft penalty is introduced, i.e.,

$$\sum_{i=1}^{N_1} (|A_1^i \odot M|_s - 1)^2 + \sum_{i=1}^{N_2} (|A_2^i \odot M|_s - 1)^2 \quad (7)$$

where A_1^i is an N_1 by N_2 coefficient matrix with 1 in the i th row and 0 for others; A_2^i is an N_1 by N_2 coefficient matrix with 1 in the i th column and 0 for other elements. The first term tends to matching each feature in the first image to a feature with the largest similarity in the second one, and the second term tends to matching each feature in the second image with a feature with the highest similarity in the first one.

Vectorizing the coefficient matrices A_1^i and A_2^i and arranging the derived column vectors, we construct the constraint coefficient matrices \hat{A}_1 and \hat{A}_2

$$\hat{A}_1(:, i) = \text{vec}(A_1^i) \quad \hat{A}_2(:, i) = \text{vec}(A_2^i).$$

Then the item (7) can be expressed as

$$\begin{aligned} & Tr \left(\left(\hat{A}_1^T \vec{M} - e_{N_1} \right)^T \left(\hat{A}_1^T \vec{M} - e_{N_1} \right) \right) \\ & + Tr \left(\left(\hat{A}_2^T \vec{M} - e_{N_2} \right)^T \left(\hat{A}_2^T \vec{M} - e_{N_2} \right) \right) \end{aligned} \quad (8)$$

where $\hat{A}_1 = e_{N_2} \otimes I_{N_1}$ is an $N_1 N_2$ by N_1 matrix, $\hat{A}_2 = I_{N_2} \otimes e_{N_1}$ is an $N_1 N_2$ by N_2 matrix, e_N is an N dimensional column vector of 1 and I_N is an N by N identity matrix.

Note that for the one-to-one correspondence, we can also impose hard constraints, i.e.,

$$\hat{A}_1^T \vec{M} = e_{N_1} \text{ or } \hat{A}_2^T \vec{M} = e_{N_2} \quad (9)$$

but these conditions may not be satisfied, since the feature extracted in one image may not have a correspondence in the other image due to the noise, occlusion or the inequality of the feature set cardinality. Thus, we adopt a soft penalty in the objective and the affine constraints are consequently removed from the formulation.

D. Inhomogeneous Pair Labeling

Since the one-to-one matching is optimized on the product graph of the two input graphs, the number of variables can be extremely large and it grows rapidly with the increase of the input vertex number. The number of features extracted depends on various factors such as the feature extractors, the complexity of surroundings, the scales searched for maximum and the size of images. Generally, the assignment variables are highly redundant. Substantive assignment variables are dispensable due to the low similarity, or, large feature distances between the involved feature pairs. We call these assignments *inhomogeneous pairs*. Rather than simply removing them, in our framework the *mismatch* information of those inhomogeneous pairs is also employed. Specifically, they are assigned as 0's, which indicate that the corresponding feature pairs will not be matched, i.e.,

$$m_{i,j} = \vec{M}_{i+(j-1) \times N_1} \Leftarrow 0 \text{ if } \{\phi_i^1, \phi_j^2\} \in \Psi \quad (10)$$

where Ψ is the set of inhomogeneous pairs. Then the *mismatch* information of those inhomogeneous pairs is also utilized to guide the solution and transduced to the remaining ones.

E. Reliable Correspondence Propagation

In the following the known correspondences including some reliable correspondences and certain number of inhomogeneous pairs are called labeled assignments or labeled feature pairs. We arrange the matching variables so that the labeled assignments are placed ahead, i.e.,

$$\vec{M}^* = [\vec{M}^l; \vec{M}^u] \quad (11)$$

where \vec{M}^l represents the assignments of the labeled feature pairs, \vec{M}^u corresponds to the assignment values of the remaining unlabeled feature pairs to be estimated. \vec{M}^* is the rearranged assignment vector.

Correspondingly, the constraint coefficient matrices \hat{A}_1 , \hat{A}_2 and the vectorized adjacency matrix \vec{S} of the similarity graph are also rearranged, so that

$$\hat{A}_1^* = [\hat{A}_1^l; \hat{A}_1^u], \hat{A}_2^* = [\hat{A}_2^l; \hat{A}_2^u], \text{ and } \vec{S}^* = [\vec{S}^l; \vec{S}^u] \quad (12)$$

where \hat{A}_1^l , \hat{A}_2^l , and \vec{S}^l are the coefficients and vectorized adjacency sub-matrix of the similarity graph for the labeled assignments, respectively; \hat{A}_1^u , \hat{A}_2^u , and \vec{S}^u are the coefficients and vectorized adjacency sub-matrix for the unlabeled assignments;

and \hat{A}_1^* , \hat{A}_2^* , and \vec{S}^* are the rearranged coefficients and vectorized similarity graph adjacency matrix.

Due to the variable rearrangement, the vertex order in the categorical product graph is also modified. The rearranged adjacency matrix W^{a*} and the corresponding Laplacian matrix L^{a*} are

$$\begin{aligned} W^{a*} &= \begin{pmatrix} W_{ll}^a & W_{lu}^a \\ W_{ul}^a & W_{uu}^a \end{pmatrix} \\ L^{a*} &= \begin{pmatrix} L_{ll}^a & L_{lu}^a \\ L_{ul}^a & L_{uu}^a \end{pmatrix}. \end{aligned} \quad (13)$$

Integrating all factors and we get the final optimization formulation for our feature matching framework:

$$\begin{aligned} \min_{\vec{M}^{u*}} & -\vec{S}^{*T} \vec{M}^* + \lambda \vec{M}^{*T} L^{a*} \vec{M}^* \\ & + \gamma \left(\text{Tr} \left(\left(\hat{A}_1^{*T} \vec{M}^* - e_{N_1} \right)^T \left(\hat{A}_1^{*T} \vec{M}^* - e_{N_1} \right) \right) \right. \\ & \left. + \text{Tr} \left(\left(\hat{A}_2^{*T} \vec{M}^* - e_{N_2} \right)^T \left(\hat{A}_2^{*T} \vec{M}^* - e_{N_2} \right) \right) \right) \\ \text{w.r.t. } & \vec{M}_i^{u*} \in \{0, 1\}, i \in \{1, 2, \dots, N_u\} \end{aligned} \quad (14)$$

where \vec{M}_i^{u*} is the i th element of \vec{M}^{u*} , N_u is the number of unlabeled feature pairs, λ and γ are coefficients controlling the balance among feature similarity, spatial coherency and one-to-one penalty.

Algorithm 1 Elicit k Correspondences. [Input: M]

- 1) Output the correspondence $m_{ij} = \{\phi_i^1, \phi_j^2\} = \arg \max_{\phi_i^1, \phi_j^2} M$.
- 2) Remove from M all potential assignments in conflict with m_{ij} .
- 3) If column or row dimension of M becomes 0 or if the output correspondence number reaches k , stop; otherwise, go back to step 1.

We relax the binary integer optimization problem to real values by discarding the constraints in (14) and the formulation is converted to an unconstrained quadratic optimization. Taking the derivative w.r.t. \vec{M}^* and substituting the (11), we obtain a closed-form relation between the labeled and unlabeled assignments:

$$\vec{M}^{u*} = C_{uu}^{-1} (B_u - C_{ul} \vec{M}^l) \quad (15)$$

where

$$\begin{aligned} C &= \begin{pmatrix} C_{ll} & C_{lu} \\ C_{ul} & C_{uu} \end{pmatrix} \\ &= \gamma \left(\hat{A}_1^* \hat{A}_1^{*T} + \hat{A}_2^* \hat{A}_2^{*T} \right) + \lambda L^{a*} \end{aligned} \quad (16)$$

and

$$\begin{aligned} B &= \begin{pmatrix} B_l \\ B_u \end{pmatrix} \\ &= \gamma \left(\hat{A}_1^* e_{N_1} + \hat{A}_2^* e_{N_2} \right) + \frac{1}{2} \vec{S}^*. \end{aligned} \quad (17)$$

F. Rearrangement and Discretizing

To get the original assignment \mathcal{M} , we first take the inverse process of the element arrangement described above and convert \vec{M}^* to \vec{M} , then reshape the derived assignment vector into the N_1 by N_2 matrix M . Since the assignment variables have been relaxed, we tried two discretization strategies: thresholding and eliciting. Setting a threshold τ for discretization is natural and it can determine the correspondence number automatically. This strategy is also suitable for the cases in which the correspondences are not required to be one-to-one. On the other hand, in case a fixed number of one-to-one correspondences are needed, we design an iterative correspondence eliciting procedure, which is displayed in Algorithm 1. Finally, the whole algorithmic process is listed in Algorithm 2.

To conclude, we deem the matches as vertices in the categorical product graph of the two input spatial graphs, and the edges then represent the geometrical relations of the matches. Similar to the semi-supervised learning algorithms, our algorithm propagates the correspondence information from the labeled to the unlabeled matching variables. The labeled matching variables are provided by two parts, i.e., reliable correspondences for matches and inhomogeneous labeling for mismatches. The connection between the labeled and those unlabeled is derived as a closed form solution in Section II-E.

III. ALGORITHMIC ANALYSIS

A. Selection of Reliable Correspondences

The accuracy of those reliable correspondences are critical for final performance. One way to obtain these reliable correspondences in the automatic matching configuration is simply to pick up a few pairs with the highest similarity scores or to set a certain threshold between the maximum correspondence and the second largest correspondence as in [17]. Fig. 4 demonstrates the matching between two human head image pair, and the algorithm in [17] with a small threshold is used for the robust feature selection. One problem is that the correspondences derived in this way may be clustered together and their guidance for the correspondence searching is, thus, limited. The work [11] proposes an Adaptive Non-Maximal Suppression (ANMS) strategy to elicit a fixed number of interest features and at the same time keep the the selected interest points spatially well distributed. Also, [26] proposes a reliable matching algorithm based on solving a low-complexity graph problem. All these algorithms can be used for the reliable correspondence selection. In this paper, we adopt the Correspondence Elicit Procedure described in Algorithm 1 and the first k correspondences produced are regarded to be reliable in the automatic matching configuration. The selection of parameter k and the algorithmic sensitivity with respect to k is discussed in the following sections.

The transductive property of our algorithm makes it easy to incorporate human interactions for the correspondence searching and a semi-supervised matching framework is naturally derived. In this work, two configurations of human interactions are used.

Exact Pairwise Correspondence Labeling: In this configuration, the users are asked to give exact correspondence labeling for the guidance of matching, and the assignments labeled by



Fig. 4. Demonstration of facial image matching. (a) SIFT feature matching proposed in [17] with threshold $t = 0.86$. (b) Better matching result by the proposed RCP algorithm, in which a small threshold $t = 0.6$ in [17] is used to elicit the robust correspondences for propagation. In both images, 60 matches are produced.

human are used as reliable correspondence priors in the feature matching process.

Obscure Correspondence Guidance: To facilitate the user labeling, we also provide an obscure matching scheme in which the user only has to describe a rough correspondence of image parts. The procedures used in the automatic matching configuration are then employed to extract reliable correspondences within the indicated corresponding areas.

Algorithm 2 Reliable Correspondence Propagation

- 1) **Graph Construction:** Construct the spatial graphs G^1 and G^2 from the feature locations and calculate the adjacency matrix for the categorical product graph using $W^a = W^{2T} \otimes W^1$. Construct the bipartite similarity graph G^{12} according pairwise feature similarity.
- 2) **Constraint Coefficient Matrix Initialization:** Initialize the constraint coefficient matrices \hat{A}_1 and \hat{A}_2 according to the cardinality of input feature sets.
- 3) **Assignment Labeling:** Initialize the corresponding assignments for those reliable pairs as 1 and set the assignment variables as 0 for those inhomogeneous pairs with low similarity values.
- 4) **Correspondence Propagation:** Rearrange the assignment variables, the adjacency matrices, the constraint coefficient matrices so that the labeled assignments are placed in front of the unlabeled variables and calculate the closed-form solution in (Section II-E).
- 5) **Rearrangement:** Take the inverse process of the arrangement in step 4 and get the correspondences using the strategies described in Section II-F.

B. Computational Complexity

The complexity of the inverse operation for an n by n matrix is $\mathcal{O}(n^{2.367})$ (the Coppersmith–Winograd algorithm), or $\mathcal{O}(n^3)$ (the Gaussian elimination algorithm), which is greater than the spectral algorithms ($\mathcal{O}(n^2)$) [16]. However, the matrix C_{uu} in our algorithm is sparse and exploiting this sparsity, the computational cost can be greatly reduced. Also, efficient parallel algorithms exist for the Gaussian elimination procedure in the computation of the sparse matrix inversion problem, and, thus, the computation time can be further shortened. Another factor affecting the computation cost is the candidate matching

variable number, which determines the dimension of the matrix C_{uu} . In our experiments, 6000 assignments with the largest similarity scores are fetched as matching candidates and the variable number can be adjusted according to the requirement of the applications. Our algorithm is much faster than the QP and SDP based algorithms and is applicable for the large scale real-world applications.

IV. APPLICATIONS AND EXPERIMENTS

In this section, our algorithm is systematically evaluated in two settings: unsupervised and semi-supervised. In the unsupervised setting, those reliable correspondences are derived automatically, while in the semi-supervised setting, the reliable correspondence priors are labeled manually. In all the experiments, the SIFT [17] descriptor is used for feature extraction and representation; the spatial graph is constructed using 10-nearest neighbors and the weights for the spatial graphs are calculated using heat kernels $K(x, y) = \exp\{-\|x - y\|^2 / \delta_o^2\}$ with parameters $\delta_o = 2^{1/2.5} \delta$ applied, where δ is the standard deviation of the feature locations. For the similarity graph, 16 nearest neighbors are used and the cosine distance is directly used as the graph weight. The coefficient λ is empirically set as 0.4 and γ is set as 0.05. In the inhomogeneous pair labeling process, we keep 6000 pairs with the top similarities as candidate matchings and others are labeled 0. The performance of our algorithm is systematically compared with the state-of-the-art feature matching algorithms, such as the spectral correspondence technique (SC) [16] and the matching algorithm used in [17] (SM), which compares the distance of the closest neighbor to that of the second-closest neighbor. We take the N_1 by N_2 pairwise similarity matrix as the input M for the Correspondence Eliciting Procedure (CE) and the matching scores are also reported. The QP and SDP based algorithms are inapplicable for comparison due to the large number of features involved. For the adjacency matrix M in the spectral correspondence technique [16], we assign a score that is linearly increasing with the cosine distance between the feature and its candidate corresponding feature to the diagonal element. Since the adjacency matrix of the categorical product graph in our algorithm represents the geometrical relations of assignments, the nondiagonal elements of M is set using the corresponding elements in W^a .

A. Automatic Feature Matching on Oxford Image Transformation Database

In this subsection, the unsupervised version of our algorithm is evaluated on the Oxford real image transformation database [1]. The Oxford database is a benchmark database for the feature descriptor evaluation. It contains eight subsets for six different geometric and photometric real image transformations, including zoom, rotation, viewpoint change, image blur, JPEG compression, and light variation. Two different scene types are involved for the case of rotation, viewpoint change, and blur: one contains homogeneous regions with distinctive edge boundaries and the other contains repeated textures of different forms, which facilitates us to analyze the effect of changing the image conditions and the scene type separately. Some images in Oxford database are demonstrated in Fig. 5. The image width and height are resized to 1/5 of the original ones and for each image,

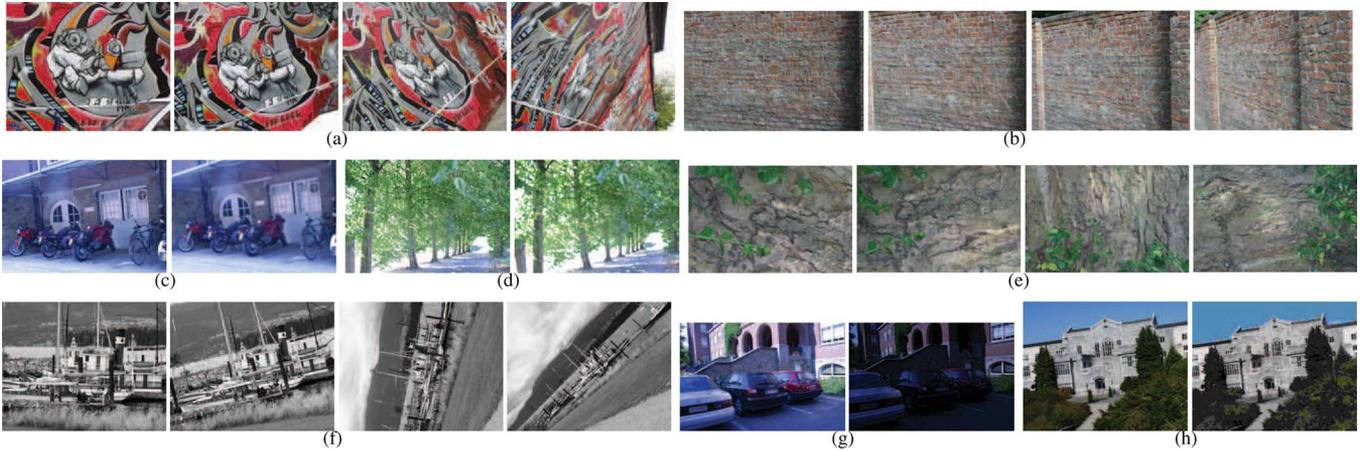


Fig. 5. Oxford real Image Transformation Database. The transformations include viewpoint variation [(a) Graffiti; (b) Wall sequence], image blur [(c) bikes and (d) trees sequence], zoom and rotation [(e) bark and (f) boat sequence], illumination change [(g) Leuven] and JPEG compression [(h) UBC].

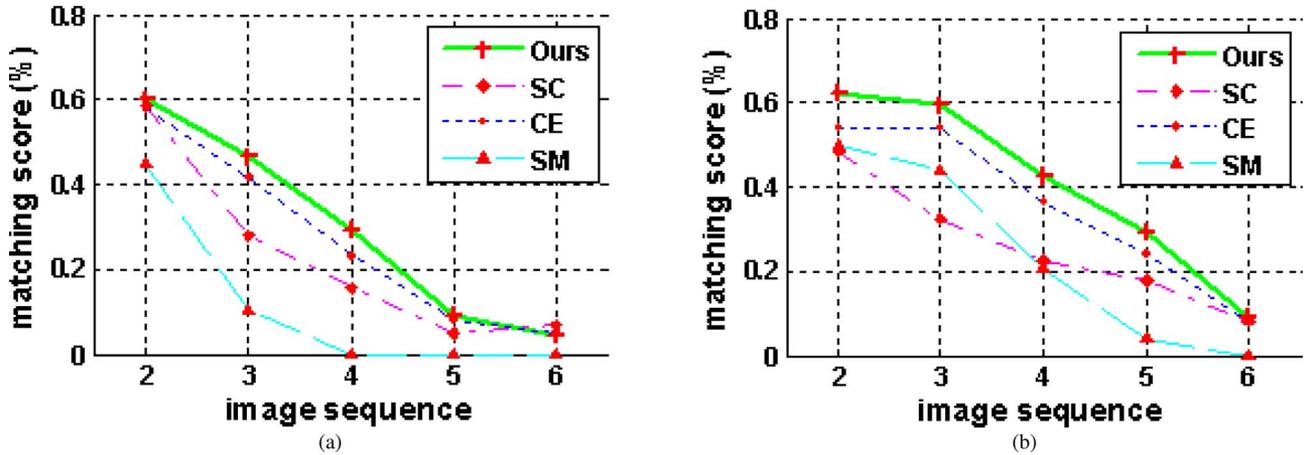


Fig. 6. Automatic feature matching score on the (a) Graffiti and (b) Wall sequence for viewpoint change. The camera varies from a fronto-parallel view to one with significant foreshortening at approximately 60 degrees to the camera. The image sequence is ordered with an increasing viewpoint variation.

100–500 SIFT descriptors are extracted. Since the homographies between the reference image and other images in each particular subset are given, we can derive the ground truth matches for the evaluation.

40–180 assignments are extracted as the reliable correspondences using Algorithm 1 in the evaluation. The matching score is calculated as the ratio between the number of correct matches and the smaller value of detected feature numbers from the image pair. For example, if 100 and 130 features are extracted from the input images, respectively, we can at most find $t_m = 100$ possible one-to-one matches and then the algorithms generate $t_m = 100$ matches in the experiment. If $c_m = 60$ correct matches are found, the matching score is $c_m/t_m = 60/100 = 60\%$ and the corresponding mismatch score is $(t_m - c_m)/t_m = 40\%$.

The detailed results are demonstrated in Figs. 6–10. It is observed that our algorithm generally reaches a higher accuracy compared with the state-of-the-art techniques and the algorithmic performance is stable over all the subsets. Although, in some situations, such as the JPEG compression the spectral technique shows an excellent performance, it is not so stable in most cases, and, thus, the matching score of the spectral

techniques is sensitive to the configuration of the input images.

Though the spectral based technique also employs geometric information as well as feature similarity in the matching process, our algorithm generally produces a better performance. The main reason is that our algorithm essentially puts different weights on the correspondences and the reliable correspondences are emphasized, while this information is ignored in other state-of-the-art feature matching algorithms.

The time consumed by different algorithms is also listed in Table I, from which we can see that the proposed RCP algorithm is comparably time-consuming especially when the feature number grows large, but the overall runtime of RCP is still acceptable in applications. For a 100 to 100 matching problem, our algorithm takes less than 30 s using the MATLAB software in a 2.5-GHZ Pentium computer.

B. Influence of Reliable Correspondence Number

In the unsupervised configuration, the performance of our algorithm relies on the accuracy of the reliable correspondences, which also deteriorates as the correspondence number increases. It is interesting and necessary to evaluate the performance of our

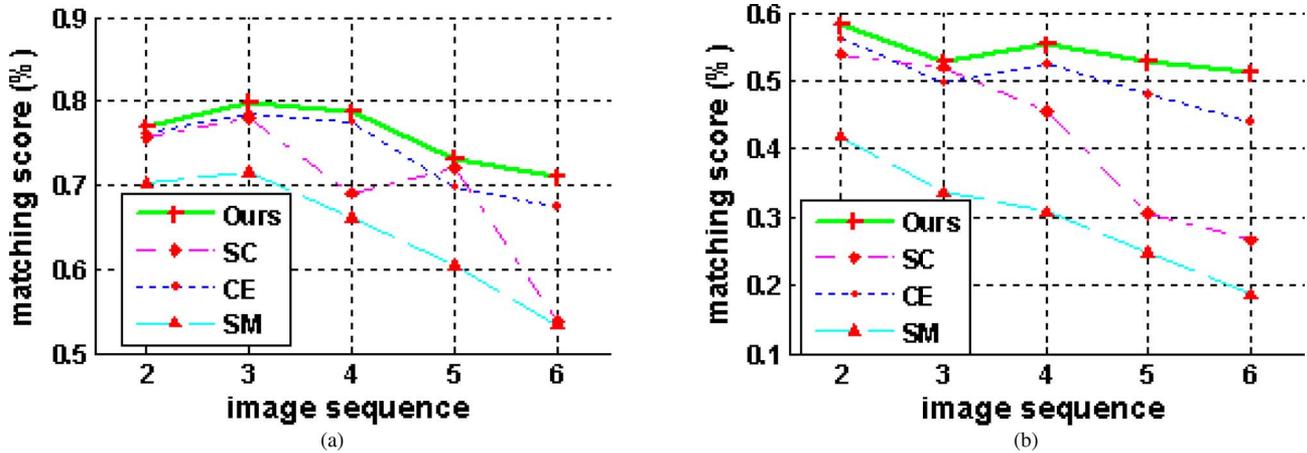


Fig. 7. Automatic feature matching score on the (a) bikes and (b) trees sequence for image blur. The blur sequence is acquired by varying the camera focus and the sequence is ordered with an increasing image blur.

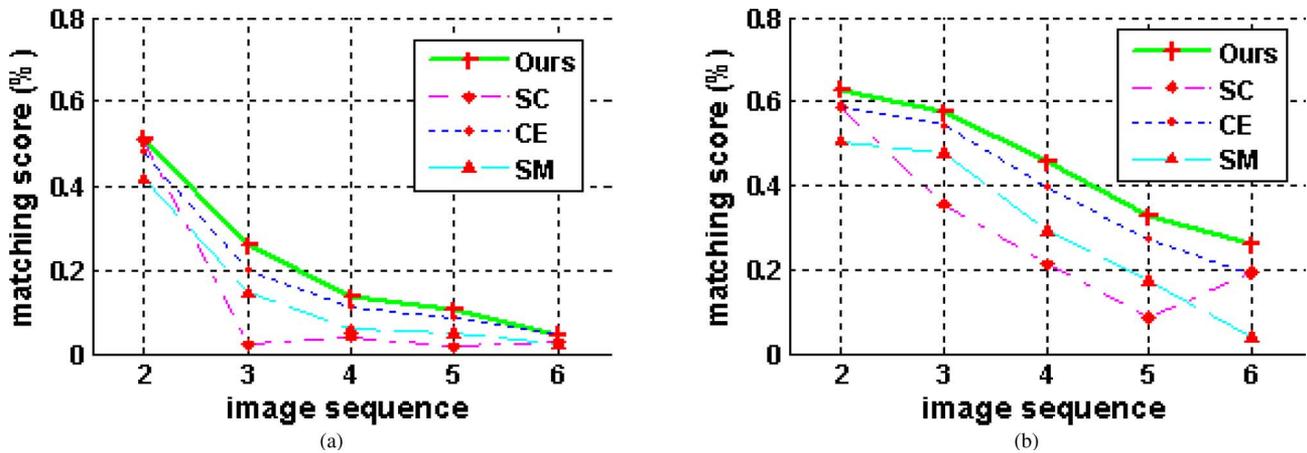


Fig. 9. Automatic feature matching score on the (a) bark and (b) boat sequence for zoom and rotation. The scale change (zoom) is acquired by varying the camera zoom and the zoom changes by about a factor of four. The image sequences are ordered with an increasing zoom and rotation.

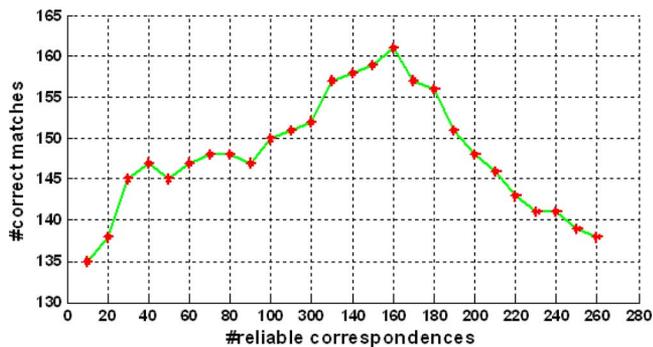


Fig. 8. Number of correct matches versus the number of automatically selected reliable correspondences on the first two images of Graffiti database.

algorithm with respect to the number of automatically selected *reliable* correspondences. Fig. 8 shows the correct matching number versus the number of reliable correspondences automatically derived. We can observe that the correct match number increases along with the increase of the reliable correspondence number within a reasonable range, and then the accuracy falls when the reliable correspondence number becomes too large to give an accurate guidance.

C. Matching Demonstration on Object Recognition Databases

In this subsection, we evaluate our algorithm on the Caltech 101 Object Recognition database [2] and ETH-80 database [3]. Four categories of images are used in this demonstration, i.e., the *motorbikes* and *face* images from Caltech 101 database as well as the *dog* and *horse* images from the ETH-80 database. Since for the objects of different types, the correspondences may not be one-to-one, a threshold of $\tau = 0.01$ is used in the discretization process, and, thus, the correspondence number is determined automatically. For comparison, the matchings with the largest k cosine distances are also plotted as baseline, where k is the number of correspondences determined by our algorithm. The matching results are demonstrated by Figs. 11–13, in which the reliable correspondences drawn by hand are marked by red stars, the obscure guidance indicated by human interaction is described by rectangles of different colors and the automatically derived reliable correspondences are plotted by small crosses. The correspondence number of the two figures within the same column is the same. Figs. 11 and 12 show the matching improvement derived by using information provided by human interaction. Fig. 13 demonstrates the improvement of matching quality in an automatic configuration. The matches are indicated by the

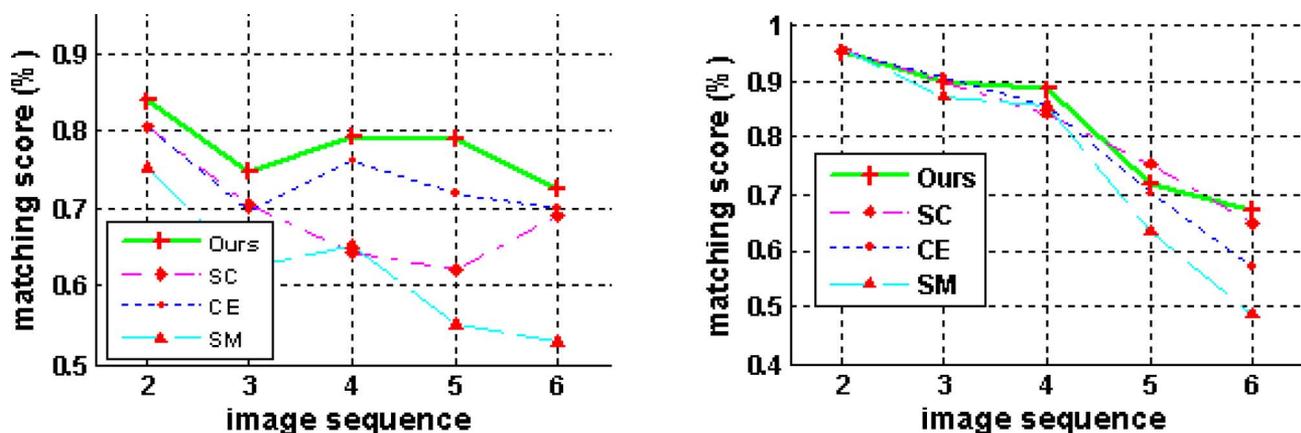


Fig. 10. Automatic feature matching score on (a) Leuven for illumination variation and (b) UBC for JPEG compression. The illumination variation sequence is introduced by decreasing the camera aperture. The JPEG sequence is generated using a standard xv image browser with the image quality parameter varying from 40% to 2%.

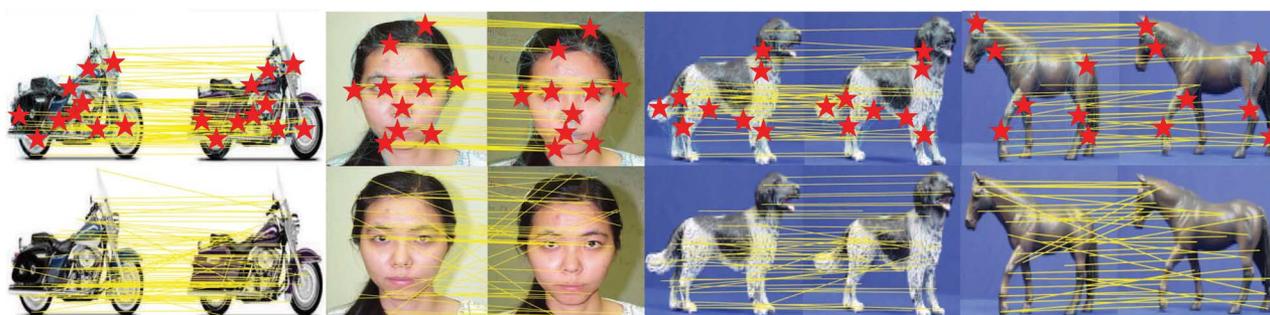


Fig. 11. Semi-supervised RCP results by manual pairwise correspondence labeling (first row) versus baseline algorithm (second row).

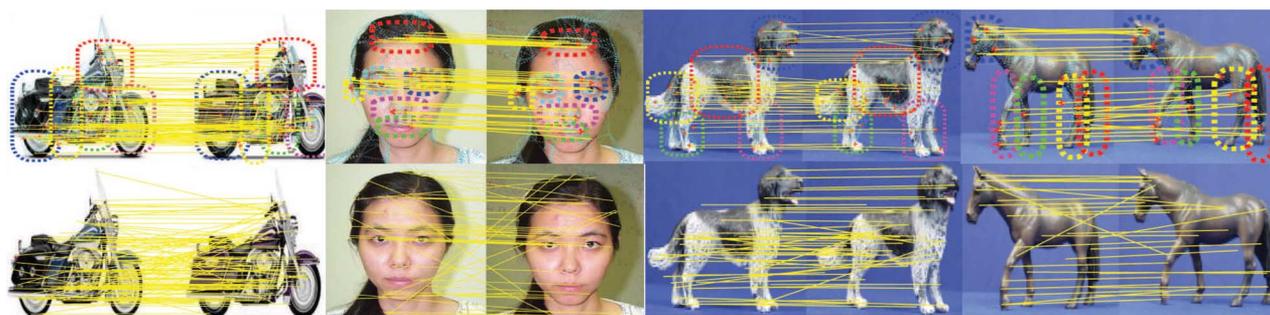


Fig. 12. Semi-supervised RCP results of obscure correspondence guidance (first row) versus baseline algorithm (second row).

TABLE I
RUNTIME (SECOND) OF RCP, SC, CE, AND SM ALGORITHMS.
THE EXPERIMENTS ARE IMPLEMENTED USING THE MATLAB
SOFTWARE IN A 2.5-GHZ PENTIUM COMPUTER

configure	RCP	SC	CE	SM
10×10	0.01	0.04	0.0002	0.0007
30×30	0.5	0.2	0.0008	0.0008
60×60	9.6	0.5	0.004	0.002
100×100	24.5	2.8	0.006	0.004
200×200	31.9	9.9	0.01	0.01

yellow line and we can observe clearly for the baseline algorithm that there exist lots of “cross lines” connecting from down-side to upside or vice versa, which indicates incorrect matches, while for the RCP algorithm the matches derived are comparably more consistent. We can also observe from the result that

the matching accuracy is boosted with the guidance of the manually labeled correspondences, and the unsupervised version of our algorithm is still superior over the baseline algorithm.

V. CONCLUSION AND FUTURE WORKS

In this paper, we proposed an efficient feature matching framework that transduces certain number of reliable correspondences to the remaining ones by utilizing both geometric coherency constraints and feature agreements. Furthermore, the framework is naturally extended to incorporate human interactions for promoting feature matching performance. Experimental results showed that our algorithm, both semi-supervised and unsupervised versions, achieves a higher matching accuracy compared to the state-of-the-art techniques. We are

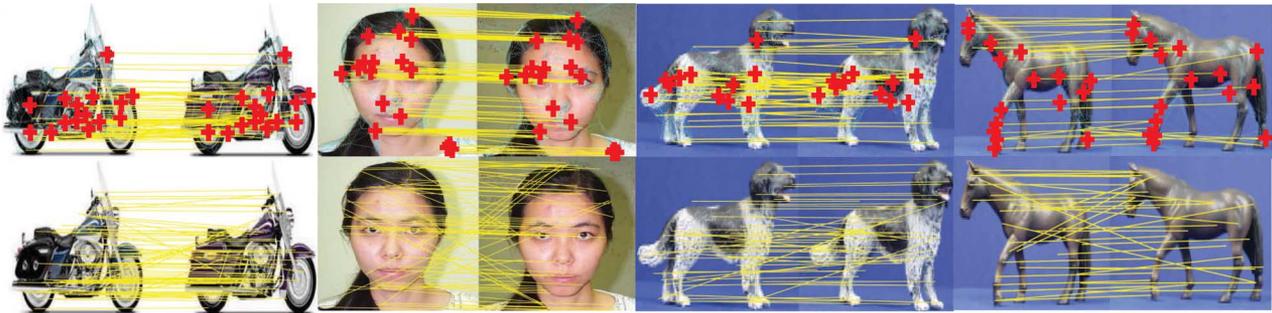


Fig. 13. Matching results of unsupervised RCP (first row) versus baseline algorithm (second row).

planning to further investigate our algorithm with other feature descriptors and explore the combination with the ANMS strategy and [26] for reliable correspondence selection.

REFERENCES

- [1] [Online]. Available: <http://www.robots.ox.ac.uk/vgg/research/affine>
- [2] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An Incremental bayesian approach tested on 101 object categories," presented at the IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR) Workshop on Generative-Model Based Vision, 2004.
- [3] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," presented at the IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 2003.
- [4] Y. Ahn, T. Yoon, and T. Schenk, "Reconstruction of 3-D object space from imagery by feature-based matching," presented at the Geoscience and Remote Sensing Symp., Jul. 2005.
- [5] C. Baillard, C. Schmid, A. Zisserman, and A. Fitzgibbon, "Automatic line matching and 3-D reconstruction of buildings from multiple views," in *Proc. ISPRS Conf. Automatic Extraction of GIS Objects From Digital Imagery*, Sep. 1999, vol. 32, pp. 69–80, Part 3-2W5.
- [6] T. Beier and S. Neely, "Feature-based image metamorphosis," in *Proc. Special Interest Group on GRAPHics and Interactive Techniques*, 1992, pp. 35–42.
- [7] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, 2003.
- [8] M. Belkin, P. Niyogi, and V. Sindhwani, "On manifold regularization," presented at the 10th Int. Workshop on Artificial Intelligence and Statistics, 2005.
- [9] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," presented at the IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 2005.
- [10] M. Brown and D. G. Lowe, "Recognising panoramas," in *Proc. IEEE Int. Conf. Computer Vision*, 2003, pp. 1218–1227.
- [11] M. Brown, R. Szeliski, and S. Winder, "Multi-image matching using multi-scale oriented patches," presented at the Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 2005.
- [12] T. Cour, P. Srinivasan, and J. Shi, "Balanced graph matching," presented at the Neural Information Processing Systems, 2007.
- [13] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," presented at the IEEE Int. Conf. Computer Vision, 2005.
- [14] K. Hariharakrishnan and D. Schonfeld, "Fast object tracking using adaptive block matching," *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 853–859, Oct. 2005.
- [15] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," presented at the European Conf. Computer Vision, 2004.
- [16] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," presented at the IEEE International Conf. Computer Vision, 2005.
- [17] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, 2003.
- [18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE International Conf. Computer Vision*, Washington, DC, 1999, vol. 2, p. 1150.
- [19] T. Luhmann and W. Tecklenburg, "3-d object reconstruction from multiple-station panorama imagery," presented at the ISPRS Working Group V/1 Panoramic Photogrammetry Workshop, Feb. 2004.
- [20] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vis.*, 2004.
- [21] C. Schellewald and C. Schnörr, "Probabilistic subgraph matching based on convex relaxation," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2005.
- [22] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [23] H. Bai and E. Hancock, "Graph matching using spectral embedding and semidefinite programming," presented at the Brit. Machine Vision Conf., 2004.
- [24] J. Kostlika, J. Cech, and R. Sara, "Feasibility boundary in dense and semi-dense stereo matching," *Comput. Vis. Pattern Recognit.*, 2007.
- [25] M. Lhuillier and L. Quan, "Match propagation for image-based modeling and rendering," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002.
- [26] M. Bujnak and R. Sara, "A robust graph-based method for the general correspondence problem demonstrated on image stitching," presented at the IEEE Int. Conf. Computer Vision, 2007.



Huan Wang received the B.Eng. degree from Zhejiang University and the M.Phil. degree from The Chinese University of Hong Kong, both in the major of information engineering. He is currently pursuing the Ph.D degree at the Computer Science Department, Yale University, New Haven, CT. His research interests include artificial intelligence, machine learning, and computer vision.



Shuicheng Yan (M'06) received the B.S. and Ph.D. degrees from the Applied Mathematics Department, School of Mathematical Sciences, Peking University, China, in 1999 and 2004, respectively.

His research interests include computer vision and machine learning. Currently, he is an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore.



Jianzhuang Liu (M'02–SM'02) received the B.E. degree from Nanjing University of Posts and Telecommunications, China, in 1983, the M.E. degree from Beijing University of Posts and Telecommunications, China, in 1987, and the Ph.D. degree from The Chinese University of Hong Kong in 1997.

From 1987 to 1994, he was a faculty member in the Department of Electronic Engineering, Xidian University, China. From August 1998 to August 2000, he was a research fellow at the School of Mechanical and Production Engineering, Nanyang Technological University, Singapore. Then he was a postdoctoral fellow in The Chinese University of Hong Kong for several years. He is now an Assistant Professor in the Department of Information Engineering, The Chinese University of Hong Kong. His research interests include computer vision, image processing, and machine learning.



Xiaoou Tang (S'93–M'96–SM'02–F'08) received the B.S. degree from the University of Science and Technology of China, Hefei, in 1990, the M.S. degree from the University of Rochester, Rochester, NY, in 1991, and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1996.

He is a Professor and the Director of Multimedia Lab, Department of Information Engineering, Chinese University of Hong Kong. He is also the group manager of the Visual Computing Group at the Microsoft Research Asia. His research interests include

computer vision, pattern recognition, and video processing.

Dr. Tang is a local chair of the IEEE International Conference on Computer Vision (ICCV) 2005, an area chair of ICCV'07, a program chair of ICCV'09, and a general chair of the ICCV International Workshop on Analysis and Modeling of Faces and Gestures 2005. He is a Guest Editor of the Special Issue on Underwater Image and Video Processing of the IEEE JOURNAL OF OCEANIC ENGINEERING and the Special Issue on Image- and Video-Based Biometrics of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He is an Associate Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.



Thomas S. Huang received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, R.O.C., and the M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge.

He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, Urbana, where he is now the William L. Everitt Distinguished Professor of Electrical and Computer Engineering, Research Professor at the Coordinated Science Laboratory, Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology, and Co-Chair of the Institutes major research theme Human Computer Intelligent Interaction. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 20 books and over 500 papers in network theory, digital filtering, image processing, and computer vision.

Dr. Huang is a Member of the National Academy of Engineering, a Foreign Member of the Chinese Academies of Engineering and Sciences, and a Fellow of the International Association of Pattern Recognition and the Optical Society of American. He received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Signal Processing Societys Technical Achievement Award in 1987 and the Society Award in 1991. He was awarded the IEEE Third Millennium Medal in 2000. Also in 2000, he received the Honda Lifetime Achievement Award for contributions to motion analysis. In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize, International Association of Pattern Recognition, and the Pan Wen-Yuan Outstanding Research Award. In 2005, he received the Okawa Prize. In 2006, he was named by IS&T and SPIE as the Electronic Imaging Scientist of the year. He is a Founding Editor of the *International Journal Computer Vision, Graphics, and Image Processing* and Editor of the *Springer Series in Information Sciences*, published by Springer Verlag.