Global Propagation of Affine Invariant Features for Robust Matching

Chunhui Cui and King Ngi Ngan, *Fellow, IEEE*

***Abstract*— Local invariant features have been successfully used in image matching to cope with viewpoint change, partial occlusion, and clutters. However, when these factors become too strong, there will be a lot of mismatches due to the limited repeatability and discriminative power of features. In this paper, we present an efficient approach to remove the false matches and propagate the correct ones for the affine invariant features which represent the state-of-the-art local invariance. First, a pair-wise affine consistency measure is proposed to evaluate the consensus of the matches of affine invariant regions. The measure takes into account both the keypoint location and the region shape, size, and orientation. Based on this measure, a geometric filter is then presented which can efficiently remove the outliers from the initial matches, and is robust to severe clutters and non-rigid deformation. To increase the correct matches, we propose a global match refinement and propagation method that simultaneously finds a optimal group of local affine transforms to relate the features in two images. The global method is capable of producing a quasi-dense set of matches even for the weakly textured surfaces that suffer strong rigid transformation or non-rigid deformation. The strong capability of the proposed method in dealing with significant viewpoint change, non-rigid deformation, and low-texture objects is demonstrated in experiments of image matching, object recognition, and image based rendering.**

***Index Terms*— Image based rendering, match propagation, mismatch rejection, nonrigid deformation, wide baseline matching.**

I. Introduction

MANY computer vision tasks rely on the establishing of adequate and accurate correspondences between images, for example, stereo vision, object recognition, image retrieval, camera self-calibration and so forth. Recently, local invariant features [12], [13] have been widely used to address this problem because of their robustness to partial occlusion and viewpoint change. Basically, the local features are first extracted independently from two images, and then characterized by some appearance descriptors, based on which the correspondences are finally established. Thanks to the intense research works done these years, the local features have been developed to be invariant not just to translation and rotation, but also to scale change [10] and affine transformation [11], [20], making matching under general viewpoint change become possible. Among various features, the *affine invariant features* [13] are particularly significant and in general provide more useful information. For example, their elliptical support regions describe different scales in the two principle directions rather than a single uniform scale indicated by the scale invariant features like SIFT [10]. The detailed evaluation and comparison on recently proposed local invariant feature detectors and descriptors can be found in [12], [13].

In spite of the success of local features, the repeatability of feature extraction and the correctness of feature matching remain an issue in the presence of severe clutters and challenging viewing conditions. Large scale and viewpoint changes considerably lower the probability of detecting consistent features in different images (features that capture the same physical surface but may appear different due to viewpoint change). Meanwhile extensive clutters may give rise to a large number of irrelevant features which disturb the matching. The situation is even worse when the regions of interest are poorly textured. In this case, the extracted features are much fewer and are more difficult to be distinguished from each other because they all look similar. The combination of these difficulties may result in a correspondence set with high percentage of mismatches. And any application based on such a matching result will probably fail. To cope with the problem, many efforts have been made to remove the outliers and increase the inliers, which is also the focus of this work.

A. Mismatch Rejection

The limited discriminative power of feature appearance may result in a large number of mismatches in the initial matching attempt. Therefore, to reject the outliers while keeping the inliers, i.e., match filtering, becomes a very important step in feature-based applications. The rejection of mismatches is typically based on the spatial geometry of the features, as opposed to the initial matching where only local appearance is taken into account. Rejection methods based on global spatial configuration assumes that all features undergo a rigid transformation, for example, the Hough clustering [10] and RANSAC [19]. These global filters, however, have two major problems: 1) they cannot deal with non-rigid deformation, and 2) they are sensitive to high number of outliers in the correspondence set.

To overcome these problems, the use of semi-local geometry information has been explored in the literature. Schmid and Mohr [16] use a fixed number of local features around a given

Manuscript received May 4, 2012; revised November 19, 2012; accepted January 22, 2013. Date of publication February 11, 2013; date of current version May 22, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kiyoharu Aizawa.

The authors are with the Department of Electronic Engineering, the Chinese University of Hong Kong, Hong Kong (e-mail: chunhui.cui@gmail.com; knngan@ee.cuhk.edu.hk).

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.2013.2246521

feature to determine its semi-local structure. A similar method has been proposed in [3], where a shape context is attached to each feature to describe the spatial configuration of its neighbors. Semi-local constraints are also used by Tuytelaars and Van Gool [20]. They propose an iterative method to reject mismatches based on homographies between matches of semi-local features. Carneiro and Jepson [3] also present an efficient pair-wise grouping method. The pair-wise relation is measured by the consistency of scale, distance and heading between a pair of scale invariant matches like SIFT. The three consistency measures are then combined together to achieve robustness to the scaling, translation and rotation.

In this paper, we present an efficient geometric filter specially designed for the *affine invariant features* which represent the state-of-the-art local invariant features [13]. We propose a pair-wise affine consistency measure to evaluate the consensus of feature matches by taking into account both the keypoint location and the size, shape and orientation information of the support regions. Based on the affine consistency measure, we iteratively remove the most unreliable matches and dynamically update the reliabilities of remaining matches. There are also other geometric filters specially designed for affine invariant features. Lazebnik *et al.* [8] propose to measure the geometric consistency of triples of matches. The local affine geometry is estimated by keypoint matches, while the consistency measure takes into account the shape and size of elliptical regions by examining the variation of major and minor axes. In comparison, our method makes full use of the information of keypoints' location and regions' size, shape and orientation to compute both affine geometry and consistency measure. Besides, compared with our pair-wise measure, the examination of triples will introduce much more computations and can only preserve large cluster. The early contraction proposed in [5] also measures the pair-wise consistency by making use of the size and shape information. However, this method is based on the coherence of region overlap and as a result can only apply to the features whose regions intersect with each other, e.g., the densely sampled regular features in their work. In comparison, our method does not demand the intersection of regions and is applicable for a general correspondence set of affine invariant features. In addition, the consistency of region orientation is implicitly imposed in our measure as an additional clue, while it is not considered by their metho. Moreover, our method efficiently integrates the neighborhood correlation into the consistency measure and enables the features' neighborhood adaptive to the shape and size of their support regions.

B. Match Propagation

Feature matches that survive the geometric filter are usually too sparse for the purpose of recognition or modeling. Especially for wide baseline case, features are far less likely to be repeatedly extracted the correctly matched due to significant viewpoint change between images. This urges the need to generate a lot more correct matches from the initial seed matches. The idea of "growing matches and surfaces" has been widely used in image matching and modeling. [15] try to propagate the matches by using the existing affine transformations to guide the search for further matches. This method is designed to save more existing features, but not to generate the new ones that have not been originally extracted. [18] develop a dense matching algorithm for multiple wide-baseline images. A sparse set of initial depth estimates is propagated to dense depth map by an inhomogeneous time diffusion process. [9] present a pixel-by-pixel greed propagation strategy. The information provided by sparse point matches is expanded in image space to obtain a regular grid of quasi-dense correspondences. Because an implicit assumption is made that the local transformation between patches is a translation, this method is applicable only in case of narrow baseline images. [7] further extend this pixel-based propagation method to wide baseline matching. They use a general affine model for the local transformation between the patches, and during the propagation adapt the affine transformation based on the second order intensity moments together with the epipolar geometry. The improvement is mainly demonstrated on scenes composed of planar surfaces. [6] propose to represent the scene by a dense set of rectangular patches. Their algorithm starts from a sparse set of matched keypoints, and repeatedly expands these to nearby pixel correspondences before using visibility constraint to filter away false matches.

The most relevant previous work is [5], where match propagation of affine invariant features is successfully applied to simultaneous object recognition and segmentation. In [4], Ferrari *et al.* propose to refine the matches of affine invariant features by maximizing the similarity function of color and intensity in the 6D affine space. Later in [5] they use the initial matches as the propagation attempts and employ the match refinement to generate more feature correspondences. The method can expand a single correct initial match to cover a smooth surface with many correct matches. This increases the discriminative power to identify the object and meanwhile suggests the approximate object boundary by the final set of matches. Besides, it is reported that the method has good robustness to scale, viewpoint, occlusion, clutter and non-rigid deformations. However, since the propagation strategy is purely appearance-based, it is best suitable for image regions that are well textured and sufficiently discriminative. This also means that it probably fails for uniform or low texture surfaces which unfortunately are ubiquitous in general scenes.

In this paper, we present a global match refinement and propagation approach by taking into account both the appearance similarity and the geometry consistency. By optimizing a global function, our method is able to simultaneously refine the whole set of initial matches, as opposed to the local method [5] where the matches are processed individually. More importantly the proposed pair-wise affine consistency is incorporated in the global function for regularization. This is extraordinarily useful when local regions are poorly textured and as a result local appearance is much less reliable for guiding the matching. Comparative experiments on image matching demonstrate that the proposed global method is superior in dealing with weakly textured surfaces and non-rigid deformation.

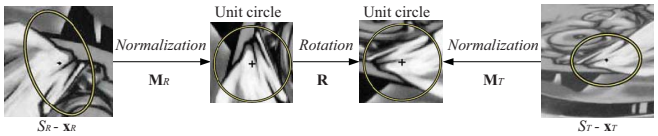


Fig. 1. Local affine transform estimated from a correspondence of affine invariant features [13].

The rest of this paper is organized as follows. Section II gives a background introduction of this work. Section III proposes the pair-wise affine consistency measure and the outlier filter. Section IV presents the global method to refine and propagate the affine invariant features. Section V discusses the application of the proposed method to image matching, object recognition and image based rendering and Section VI concludes the paper.

II. Background

In this section, we briefly describe some background knowledge which forms the basis of this work.

A. Affine Invariant Features

An affine invariant feature typically has a support region with elliptical shape [13]. It can be represented by $\mathbf{f} = [\mathbf{x}, a, b, o, \theta, \mathbf{v}]$, where $\mathbf{x}$ is the image coordinates of the feature's keypoint, a and b are the lengths of the semi-major and semi-minor axes of the feature's elliptical region, o indicates the orientation of the major axis, θ represents the region's dominant orientation which for instance can be estimated by gradient histogram [10], and $\mathbf{v}$ is the feature descriptor that summarizes the region appearance. After extracting the features, feature matching is applied to establish the correspondence set $\mathbf{\Omega}$ that associates each feature $\mathbf{f}_R$ extracted from the reference image I_R with a feature $\mathbf{f}_T$ detected in the target image I_T. In the initial stage, correspondences of features are typically established based on some similarity measure of their descriptors.

B. Local Affine Transform

With a correspondence of affine invariant features, one can estimate the local affine transform that relates the two features [13]. Let S_R and S_T denote the support regions of two matched features $\mathbf{f}_R$ and $\mathbf{f}_T$, respectively. The two regions can be centered on $(0, 0)$ by $S_R - \mathbf{x}_R$ and $S_T - \mathbf{x}_T$, where $\mathbf{x}_R$ and $\mathbf{x}_T$ are the coordinates of the corresponding keypoints. Since (a_R, b_R, o_R) and (a_T, b_T, o_T) are available, we then can normalize the two elliptical regions into unit circles by the affine transforms $\mathbf{M}_R$ and $\mathbf{M}_T$, respectively. As shown in Fig. 1, the two unit circles are now related by a pure rotation $\mathbf{R}$ which can be determined by the dominant orientations of the two features, i.e., θ_R and θ_T. Therefore, the two regions S_R and S_T are related by (1). In homogeneous coordinates, they are actually related by an affine transform $\mathbf{Af}$ (2) which can be estimated once we know $\mathbf{f}_R$ corresponds to $\mathbf{f}_T$

$$S_T - \mathbf{x}_T = \mathbf{M}_T^{-1}\mathbf{R}\mathbf{M}_R(S_R - \mathbf{x}_R) \tag{1}$$

$$\mathbf{Af} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}, \quad \text{where } \mathbf{A} = \mathbf{M}_T^{-1}\mathbf{R}\mathbf{M}_R, \ \mathbf{t} = \mathbf{x}_T - \mathbf{A}\mathbf{x}_R \tag{2}$$

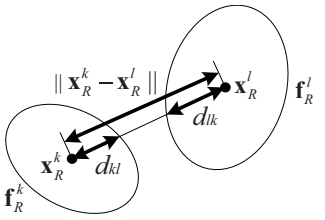


Fig. 2. Normalized spatial distance between two affine invariant features.

III. Affine Consistency and Mismatch Filter

Though the similarity measure based on feature descriptor is widely used in feature matching [12], it only relies on the region appearance that may not be sufficiently discriminative to ensure correct matches. This problem becomes more of an issue for surfaces with low texture or homogeneous texture, and probably results in a lot of false matches. In this section, we propose an efficient geometric filter to remove the mismatches from the initial correspondence set by making use of the remaining information provided by the affine invariant feature, i.e., $(\mathbf{x}, a, b, o, \theta)$. We first propose the *pair-wise affine consistency* by taking into account both the neighborhood correlation between features and the consistency of local affine geometry between matches. We then present an iterative algorithm to efficiently reject the mismatches based on the affine consistency.

A. Pair-Wise Affine Consistency

An affine transform is sufficient to locally model the image distortion arising from viewpoint changes [13]. Suppose that two neighboring features are located on the same physical surface that is smooth and approximately planar at the local scale. The two features' support regions will undergo very similar affine transforms when the viewpoint changes. This holds for deformed objects as well because generally the affine transform varies slowly and smoothly within the physical surface. Thus such two features are called *affine consistent*, and they will support each other to survive the proposed geometric filter. We believe that the more supports a match has, the more reliable it is.

Let $(\mathbf{f}_R^k, \mathbf{f}_T^k) \in \mathbf{\Omega}$ $(k = 1, \ldots, N)$ be one of the N matches in the correspondence set $\mathbf{\Omega}$. Let $\mathbf{Af}^k$ denote the affine transform that relates their support regions S_R^k and S_T^k. Let $(\mathbf{f}_R^l, \mathbf{f}_T^l) \in \mathbf{\Omega}$ $(l \neq k)$ be another match whose support regions are S_R^l and S_T^l. The pair-wise affine consistency measure $AC(k, l)$ between the two matches indexed by k and l is defined as

$$AC(k, l) = \exp(-dis(\mathbf{f}_R^k, \mathbf{f}_R^l)^2/\delta) \cdot \frac{S_T^l \cap \mathbf{Af}^k S_R^l}{S_T^l \cup \mathbf{Af}^k S_R^l} \tag{3}$$

The first term of $AC(k, l)$ measures the neighborhood correlation between the two features $\mathbf{f}_R^k$ and $\mathbf{f}_R^l$ in the reference image. Because the features represent image regions, their spatial distance should depend on not only the location of their keypoints, but also the size and shape of their support regions. Based on this consideration, the normalized spatial distance between two features $\mathbf{f}_R^k$ and $\mathbf{f}_R^l$ is defined in (4) and is illustrated in Fig. 2

$$dis(\mathbf{f}_R^k, \mathbf{f}_R^l) = \|\mathbf{x}_R^k - \mathbf{x}_R^l\|/(d^{kl} + d^{lk}). \tag{4}$$

Here d^{kl} denotes the distance from the keypoint $\mathbf{x}_R^k$ to the intersection of the elliptical region and the ray from $\mathbf{x}_R^k$ and through $\mathbf{x}_R^l$, as illustrated in Fig. 2. It can be easily computed by (5), where $\phi^{kl} = ori(\mathbf{x}_R^l - \mathbf{x}_R^k) - o_R^k$ and $ori(\cdot)$ is the vector orientation. d^{lk} is similarly defined

$$d^{kl} = \sqrt{(a_R^k \cos\phi^{kl})^2 + (b_R^k \sin\phi^{kl})^2}. \tag{5}$$

As we can see in the definition of $dis(\mathbf{f}_R^k, \mathbf{f}_R^l)$, the features' keypoint distance is normalized according to their regions' size and shape. If $dis(\mathbf{f}_R^k, \mathbf{f}_R^l) = 1$, the support region of $\mathbf{f}_R^k$ will be tangential to that of $\mathbf{f}_R^l$, and when $dis(\mathbf{f}_R^k, \mathbf{f}_R^l) < 1$, the two regions will have overlap. Simply by thresholding the normalized distance, we can determine the neighboring features in a way adaptive to the size and shape of their support regions. In general, neighboring features with small $dis(\mathbf{f}_R^k, \mathbf{f}_R^l)$ (e.g. smaller than 1) will have high probability to undergo similar local affine transform when the viewpoint changes. In (3), the inverse exponential of $dis(\mathbf{f}_R^k, \mathbf{f}_R^l)$ is reasonably used for measuring to what degree the two reference features $\mathbf{f}_R^k$ and $\mathbf{f}_R^l$ are spatially correlated and accordingly how reliable the two matches $(\mathbf{f}_R^k, \mathbf{f}_T^k)$ and $(\mathbf{f}_R^l, \mathbf{f}_T^l)$ can support each other to pass the geometric filter.

The second term of $AC(k,l)$ is to measure the consistency of the local affine geometry estimated from the two matches. Ideally, if the two reference features $\mathbf{f}_R^k$ and $\mathbf{f}_R^l$ undergo the same affine transform, we have $\mathbf{A}\mathbf{f}^k S_R^l = \mathbf{A}\mathbf{f}^l S_R^l = S_T^l$. In practice, however, the two regions $\mathbf{A}\mathbf{f}^k S_R^l$ and S_T^l will differ from each other. Thus the difference of the two regions is employed to measure the inconsistency of the affine transforms $\mathbf{A}\mathbf{f}^k$ and $\mathbf{A}\mathbf{f}^l$. Specifically, the difference between $\mathbf{A}\mathbf{f}^k S_R^l$ and S_T^l is quantified by their overlap in image area. In (3), $S_T^l \cap \mathbf{A}\mathbf{f}^k S_R^l$ is the intersection of the two regions, which is then normalized by their union $S_T^l \cup \mathbf{A}\mathbf{f}^k S_R^l$. The overall AC value is ranged from 0 to 1. A large $AC(k,l)$ value indicates that the two matches $(\mathbf{f}_R^k, \mathbf{f}_T^k)$ and $(\mathbf{f}_R^l, \mathbf{f}_T^l)$ are not only spatially correlated but also affine consistent, and hence are very likely to be a pair of correct matches.

B. Mismatch Filter Based on Affine Consistency

Given the correspondence set Ω with N matches, we first build the $N \times N$ **AC** matrix whose entries are $AC(k,l)$ $(k, l \in [1, \ldots, N])$, where $AC(k,l) = 0$ if $k = l$. The AC score for a match indexed by k is calculated by

$$\mathrm{AC}_k = \sum_l AC(k,l) \tag{6}$$

In general, $AC(k,l) \neq AC(l,k)$ due to the asymmetry of the second term in (3). However, in practice the two values are very close, thus we empirically assume symmetric **AC** matrix to reduce half of the computations. To further save the computations, we only evaluate the AC measures for neighboring matches with $dis(\mathbf{f}_R^k, \mathbf{f}_R^l) < th_1$. As the normalized distance $dis(\mathbf{f}_R^k, \mathbf{f}_R^l)$ is already incorporated in (3), AC measures of far away match pairs will have very small values and contribute little to the AC score anyway. In this paper, we empirically set th_1 to 2 and find it works well throughout the experiments. One may consider increasing th_1 only when the initial matches are too sparse.

Next, we iteratively remove the inconsistent matches from Ω and meanwhile update the **AC** matrix. The algorithm is outlined as follows.

1) Compute the AC scores for all matches in the current set Ω by (6) based on the current **AC** matrix;
2) Remove from Ω all the matches $(\mathbf{f}_R^k, \mathbf{f}_T^k)$ whose $\mathrm{AC}_k = 0$, and update the **AC** matrix by deleting the corresponding rows and columns;
3) For the remaining matches in Ω, find the one with the smallest AC score, i.e., $\mathrm{AC}_{\min}$.
4) If $\mathrm{AC}_{\min} > th_2$,[1] stop. Otherwise, remove the match and update the **AC** matrix accordingly, then go to step 1.

Matches with no support are directly rejected as mismatches because isolated correct matches are very rare in practice. Furthermore, the worst match in terms of AC score is the most probable mismatch in the remaining correspondence set. Removing this match from the **AC** matrix will largely reduce the AC scores of other nearby mismatches with similar wrong affine transforms, but has little influence on the AC scores of nearby correct matches. As a result, the nearby mismatches, as lose the support from the worst match, are more likely to be filtered out in following iterations.

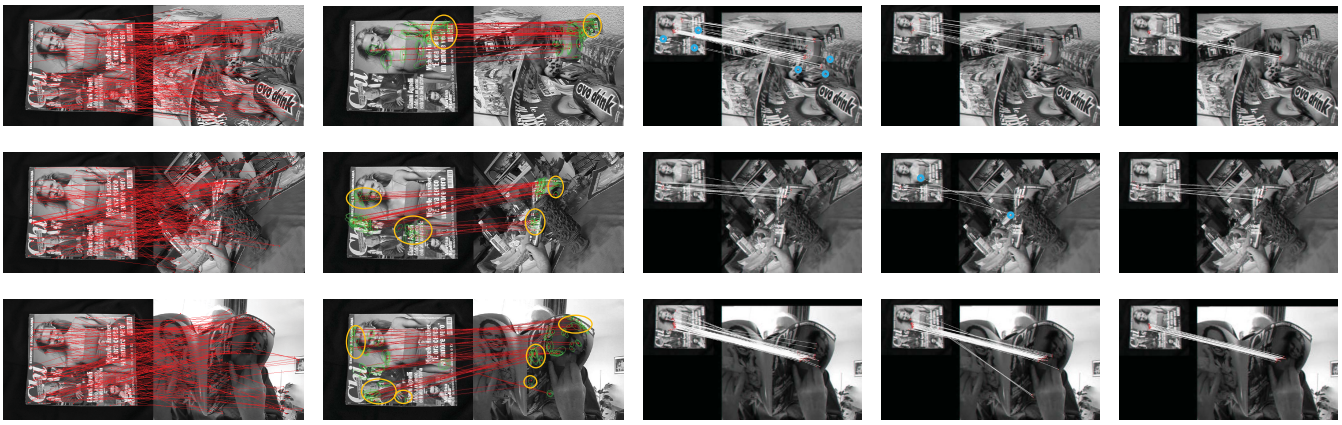
C. Performance on Image Pairs With Nonrigid Deformation

The affine consistency filter is tested on image pairs that present significant non-rigid deformation and clutters, which is much more challenging than the case of rigid transform. We use Affine Harris and Affine Hessian detectors [11] to extract the affine invariant features which are described by the standard SIFT descriptor [10]. The similarity of features is measured by the Euclidean distance of their SIFT descriptors. The strategy of *nearest neighbor distance ratio* [12] is adopted to generate the initial matches. We then apply the proposed filter to reject the outliers, resulting in the final correspondence set.

The results of feature matching for *Michelle* image set [5] are visually presented in Fig. 3, where three tests with the same reference image but different target images are shown in the three rows, respectively. Fig. 3(a) presents the initial matching results and Fig. 3(b) shows the the remaining matches that pass the proposed filter. The red lines in Fig. 3(a) and (b) indicate the correspondences and the green ellipses show the features' support regions. As we can see, all the mismatches are successfully removed by the proposed filter despite the significant non-rigid deformation between images. Also note that there exist a large number of mismatches in the initial correspondence set, which demonstrates the strong power of the proposed filter in dealing with extensive clutters.

For comparison, in Fig. 3(c)–(e), we also present the matching results of Phase and SIFT features filtered by pairwise grouping, semi-local method and hough transform, respectively, which are tested on the same image set and reported in [3]. Several obvious mismatches by pairwise

[1] th_2 is empirically set to 0.1 throughout our experiments.



(a) (b) (c) (d) (e)

Fig. 3. Apply the proposed filter to the matching results of Affine Harris and Hessian features, on the image set *Michelle* [5]. (a) Initial matches (a lot of mismatches). (b) Matches after applying the proposed filter (all correct). (c)–(e) Matching results by pairwise grouping, semi-local method, and Hough transform reported in [3], respectively, which are tested on the same image set.

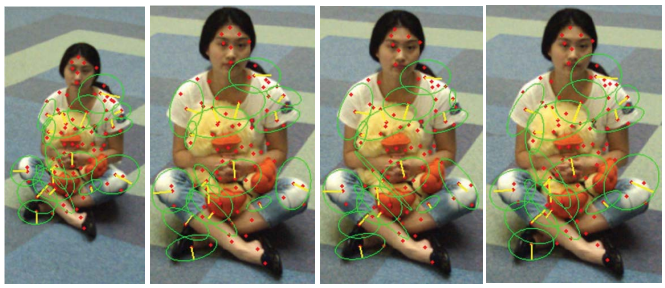
TABLE I
NUMBER OF CORRECT MATCHES BY DIFFERENT METHODS

Test	Proposed (initial)	Pairwise	Semi-Local	Hough
1	**55** (286)	39	17	10
2	**38** (172)	13	7	10
3	48 (219)	45	35	20

grouping and semi-local method are marked by blue circles in Fig. 3(c) and (d). Table I summarizes the number of correct matches found by different methods. The number in the parenthesis in the "Proposed" column indicates the amount of initial matches found by Affine Harris and Hessian features. Note that one-to-one correspondence is imposed in our method to remove the repetitive matches,[2] while Carneiro *et al.* [3] allow for many-to-many mapping in their test. Therefore, there are a few repetitive matches in Fig. 3(c)–(e) that are counted in Table I columns "Pairwise," "Semi-local" and "Hough." Compared with the methods proposed in [3], the Affine Harris and Hessian features combined with our filter can produce more correct matches, especially for test 1 and 2 as shown in Table I. More importantly, matches found by our method are able to cover more regions of the object, as marked by orange ellipses in Fig. 3(b). It is worth noting that besides the absolute amount of matches, the region coverage of matched features is also a very important clue that can be used to improve the precision of object recognition. In addition, the runtime of the proposed filter (implemented in non-optimized MATLAB code) measured on a Core Duo T2400 1.83 GHz windows laptop is around 1.68 s, 0.82 s and 1.02 s when applied to Fig. 3 test 1, 2, and 3, respectively, which is very efficient.

IV. GLOBAL MATCH REFINEMENT AND PROPAGATION

The method of match refinement and propagation [5] has been successfully applied to recognize and segment the objects. In practice, however, we found that this method works well for highly textured regions, but usually fails to establish correct correspondences for regions with low texture or homogeneous texture.[3] Some examples are shown in Fig. 4, where the red dots in (a), (b) and (c) indicate the keypoints detected in the reference image, their initial correspondences found in the target image and the refined results obtained by Ferrari's method [4], respectively. The green ellipses show the support regions of the features that are not accurately matched by the individual local refinement. The yellow lines indicate the regions' dominant gradient orientations. As we can see, most of these regions are not well textured. As a consequence, the local appearance alone is not sufficiently powerful to guide the correct refinement. In this section, we propose a global method to refine and propagate the affine invariant features by incorporating the local appearance with the local affine geometry. The improved method can successfully handle the weakly textured regions that occur frequently in general scenes.



(a) (b) (c) (d)

Fig. 4. Some matching examples where the global refinement method outperforms the local method. (a) Features in the reference image. Corresponding features in the target image. (b) Initial correspondences. (c) Results by local refinement [4]. (d) Results by global refinement.

A. Global Function for Match Refinement

The proposed global match refinement is based on the observation that the surface orientations change smoothly except

[2]We only keep the match with the smallest descriptor distance.

[3]This is also pointed out in [5].

for the surface discontinuity. This inspires us to impose the smoothness constraint on the local affine transforms of neighboring matches. The original match refinement method [4] tries to find the best affine transform for each match individually. We now attempt to simultaneously find the optimal set of affine transforms for all the matches by maximizing the global function defined in (7), where $(\mathbf{f}_R^k, \mathbf{f}_T^k) \in \Omega$ $(k = 1, \ldots, N)$. The three terms in (7) are described in the following

$$F(\{\mathbf{Af}^k\}) = \sum_{k=1}^{N} \frac{1 + \mathrm{NCC}(S_R^k, \mathbf{Af}^k S_R^k)}{2} + \lambda_1 \sum_{k=1}^{N} \exp\left(-\frac{\overline{\mathrm{DL}}(S_R^k, \mathbf{Af}^k S_R^k)}{\gamma}\right) + \lambda_2 \sum_{k=1}^{N} \sum_{l \in \Psi^k} w^l AC(k,l). \quad (7)$$

The first and the second terms are defined similarly as in [4]. They are used to measure the appearance similarity of the two matched regions S_R^k in I_R and $S_T^k = \mathbf{Af}^k S_R^k$ in I_T, which are related by the affine transform $\mathbf{Af}^k$. The first term measures the intensity similarity and the second term measures the color similarity. Together they are called the data term. NCC is the normalized cross-correlation between the regions' intensity patterns, and is normalized to [0, 1] in (7). $\overline{\mathrm{DL}}$ is the average pixel-wise Euclidean distance in the CIE-L*a*b* color space. The three color bands are normalized independently to achieve the illumination invariance to some extent. The equivalence between Euclidean and perceptual distances holds for small distances only, while the larger distance only indicates that the colors are perceptually different. By taking into account this fact we choose the exponential measure ranging from 0 to 1 for the color term.

The major improvement of the global method lies in the introduction of the third term in (7), i.e., the smoothness term. In case of a smooth region for which the local appearance is not discriminative enough to guide the correct match refinement, regularization is necessary and can be achieved by further maximizing the affine consistency between neighboring features. To this end, the proposed affine consistency measure $AC(k,l)$ is employed to regularize the set of affine transforms $\{\mathbf{Af}^k\}$ in the global function. And the normalized spatial distance $dis(\mathbf{f}_R^k, \mathbf{f}_R^l)$ can be efficiently used to determine the neighboring features in the image domain.

However, choosing all the neighboring features for regularization may result in the over-smooth problem just as in the dense stereo. Two features around the depth discontinuity may be close in the image domain, but actually belong to different physical surfaces. Thus their corresponding affine transforms from one image to another can be completely different and uncorrelated. As smoothing across depth discontinuity is highly undesired, we need to carefully define the neighborhood system for regularization. In this paper, the neighborhood of a match $(\mathbf{f}_R^k, \mathbf{f}_T^k) \in \Omega$ is described by its affinity set Ψ^k defined in (8). Basically, Ψ^k is a sub-set of the neighboring matches of $(\mathbf{f}_R^k, \mathbf{f}_T^k)$ with $dis(\mathbf{f}_R^k, \mathbf{f}_R^l) < th_1$. By further imposing the constraint $AC(k,l) > th_2$, the matches in Ψ^k should be associated with the affine transforms that are very similar to $\mathbf{Af}^k$, and hence they are called the affinities of $(\mathbf{f}_R^k, \mathbf{f}_T^k)$. In a word, only the neighboring matches with similar affine transforms are used for regularization. If two matches have quite different affine transforms, they are probably located on surfaces with different orientations or depths

$$\Psi^k = \{(\mathbf{f}_R^l, \mathbf{f}_T^l) \in \Omega (l \neq k) | dis(\mathbf{f}_R^k, \mathbf{f}_R^l) < th_1,\ AC(k,l) > th_2\}. \quad (8)$$

In order to speed up the optimization, we simply threshold the AC measure instead of using some cluster algorithms. Besides, we can further limit the size of Ψ^k, since a small number of reliable affinities are sufficient for the purpose of regularization. So actually the smoothness term for a match $(\mathbf{f}_R^k, \mathbf{f}_T^k)$ is composed by the weighted sum of the AC measures of its affinities. The strategy of choosing the weights w^l will be discussed in the next sub-section. Here note that they are normalized such that $\sum_{l \in \Psi^k} w^l = 1$. Thus the smoothness term ranges in [0, 1] as well. Finally, the two parameters λ_1 and λ_2 in (7) are empirically set to 2 and 1, respectively, such that the weights for intensity, color and smoothness terms are 1:2:1.

B. Implementation of the Global Optimization

The match refinement is now an expensive global optimization problem over a large set of affine transforms $\{\mathbf{Af}^k\}$ $(k = 1, \ldots, N)$. To make this problem tractable, we decompose the global optimization into the iterations of sequential maximization problems, each of which can be formulated as the maximization of the function $f(\mathbf{Af}^k)$ defined in (9) over the 6D space of a single affine transform $\mathbf{Af}^k$, with the affinities' transforms $\{\mathbf{Af}^l \mid (\mathbf{f}_R^l, \mathbf{f}_T^l) \in \Psi^k\}$ fixed. The value of $f(\mathbf{Af}^k)$ (ranging from 0 to 4) provides a combined evaluation of the goodness of a match $(\mathbf{f}_R^k, \mathbf{f}_T^k)$ in terms of both appearance similarity and geometry consistency. Apparently, the weight w^l in (9) should reflect the confidence of using the affinity $(\mathbf{f}_R^l, \mathbf{f}_T^l)$ for regularization. Thus, a natural choice of w^l in the current iteration is the goodness of match $(\mathbf{f}_R^l, \mathbf{f}_T^l)$, i.e., $f(\mathbf{Af}^l)$ estimated in the last iteration. In the first iteration the weights are initially set to the data term of $f(\mathbf{Af}^l)$ only, because the smoothness term is not available yet. Then the weights are updated according to the f values every iteration

$$f(\mathbf{Af}^k) = \frac{1 + \mathrm{NCC}(S_R^k, \mathbf{Af}^k S_R^k)}{2} + \lambda_1 \exp\left(-\frac{\overline{\mathrm{DL}}(S_R^k, \mathbf{Af}^k S_R^k)}{\gamma}\right) + \lambda_2 \sum_{l \in \Psi^k} w^l AC(k,l). \quad (9)$$

Now the problem is how to maximize $f(\mathbf{Af}^k)$ over the 6D affine space $(t_x,\ t_y,\ s_x,\ s_y,\ \theta,\ h)$, where (t_x, t_y) is the 2D translation, (s_x, s_y) are the scales in x and y directions, and θ and h are the rotation and shear, respectively. Though an additional smoothness term is introduced to the $f(\mathbf{Af}^k)$ function, its behavior over the affine space is similar to

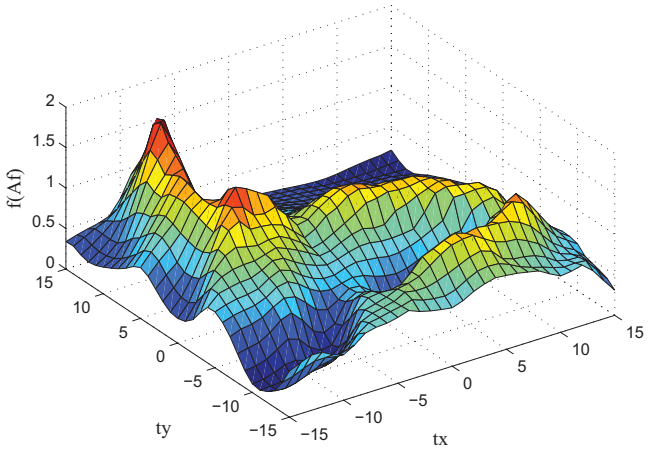


Fig. 5. Typical case of $f(\mathbf{Af}^k)$ over the space of (t_x, t_y) with fixed (s_x, s_y, θ, h), where the gradient descent algorithm starting from $(0, 0)$ fails to reach the global maxima at $(-8, 12)$.

the similarity function [4]. Fig. 5 shows a typical case of $f(\mathbf{Af}^k)$ over the space of (t_x, t_y) with (s_x, s_y, θ, h) fixed. We can see that the non-convex surface presents frequent and diverse foldings, which is the reason why Ferrari *et al.* [4] proposed to search the affine space step by step instead of using the gradient ascend.[4] In this paper, we employ a step-wise searching algorithm to do the sequential maximization. In an iteration, transforms in $\mathbf{Af}^k$ are sequentially updated for maximization of $F(\mathbf{Af}^k)$. For each $\mathbf{Af}^k$ only one step is made in the 6D affine space to approach the maxima of $f(\mathbf{Af}^k)$, that is, we only update one of the six parameters which brings the greatest ascent to $f(\mathbf{Af}^k)$. Note that it is unnecessary to fully maximize $\mathbf{Af}^k$ in one iteration because the affinities of $\mathbf{Af}^k$ may also change through iterations. In practice, we find that this step-wise algorithm performs much better than the gradient descent and generally produces satisfactory results.

The order of the sequential maximization of $f(\mathbf{Af}^k)$ $(k = 1, \ldots, N)$ may play a crucial role in the behavior of the global optimization due to the imposed smoothness term. Matches with small f values should have higher priority, because they are more likely to be inaccurate matches that badly need regularization. On the other hand, matches with high f values are probably correct and accurate, and as a result can be more reliably used to regularize other features. Therefore the matches whose affinities have high f values should also be given preference in the sequential maximization. Based on these considerations, the priority of a match in the sequential maximization is quantified by the mean f value of its affinities with respect to its own f value, as defined in (10). At the end of an iteration, the matches' priorities are updated according to the current f values, and a new order is determined for the sequential maximization in the next iteration. Besides, in order to guarantee and accelerate the convergence, we stop examining a match $(\mathbf{f}_R^k, \mathbf{f}_T^k)$ in following iterations if its associated transform $\mathbf{Af}^k$ makes no change in one maximization attempt. Such a match is called a stable match

$$priority(\mathbf{Af}^k) = \overline{f(\mathbf{Af}^l)}/f(\mathbf{Af}^k) \ (l \in \Psi^k) \qquad (10)$$

Fig. 6(a) and (b) present the performance comparison of the global optimization with and without the priority update. The results are obtained by applying the global refinement to the 87 initial matches shown in Fig. 4(a) and (b). From Fig. 6(a) and (b), we can see that a reasonable order of sequential maximization can benefit the global optimization in terms of both the convergence speed and the convergence value. One may also note that the differences are not that obvious, which means that the algorithm is not very sensitive to the sequential order. In Fig. 4(c) and (d), we select some features to visually compare the performances of the local and the global refinement. We can clearly observe the improvement achieved by the global method in terms of the accuracy of features' keypoints and support regions.

C. Global Match Propagation

Match propagation aims to generate more feature correspondences from the initial seed matches. Let Θ be the set of seed matches and Γ_R be a set of newly added features in the reference image I_R. Recall that each match $(\mathbf{f}_R^k, \mathbf{f}_T^k) \in \Theta$ is associated with an affine transform $\mathbf{Af}^k$. Thus, [5] proposed to choose for each new feature $\mathbf{f}_R^n \in \Gamma_R$ the best affine transform $\mathbf{Af}^k$ from the seed matches Θ in terms of the similarity function, which is called the best propagation attempt and is used to generate the initial correspondence of $\mathbf{f}_R^n$, i.e., $\mathbf{f}_T^n \in \Gamma_T$ in the target image I_T with $S_T^n = \mathbf{Af}^k S_R^n$. Then this initial match $(\mathbf{f}_R^n, \mathbf{f}_T^n)$ is further refined to achieve better accuracy.

In this paper, the new features are uniformly sampled within the object in the reference image. They have circular support regions of radius r and are spaced by r so that they can densely cover the object. The odd and even rows offset one other by r as well, as shown in Fig. 8. Here, the choice of the sampling parameter r trades the precision for computational cost. It could be adaptively selected according to the scene complexity or simply specified by the users.

As mentioned before, smooth regions are ubiquitous for general objects, for which the local refinement method usually fails to produce accurate matches because the local appearance is not sufficiently discriminative. To address the problem, we employ the proposed measure $f(\mathbf{Af}^k)$ (9) to select the best propagation attempt instead of using the purely appearance-based similarity measure [4]. Then, to refine the initial matches of new features, we apply the proposed global refinement method by taking into account both the appearance likelihood and geometry consistency. The detailed algorithm of global propagation is described as follows:

1) Initialization: For each new feature $\mathbf{f}_R^n \in \Gamma_R$, initialize its match $\mathbf{f}_T^n \in \Gamma_T$ by following three steps:

1) Find the nearby seed matches as the candidates. Specifically the candidate set is defined as

$$\Phi^n = \left\{(\mathbf{f}_R^k, \mathbf{f}_T^k) \in \Theta \mid dis(\mathbf{f}_R^k, \mathbf{f}_R^n) < th_1\right\} \qquad (11)$$

[4]Actually, we have tried an inverse compositional algorithm (a gradient ascend implementation) [1] to do the maximization, but unfortunately it frequently gets stuck in undesired local maxima.

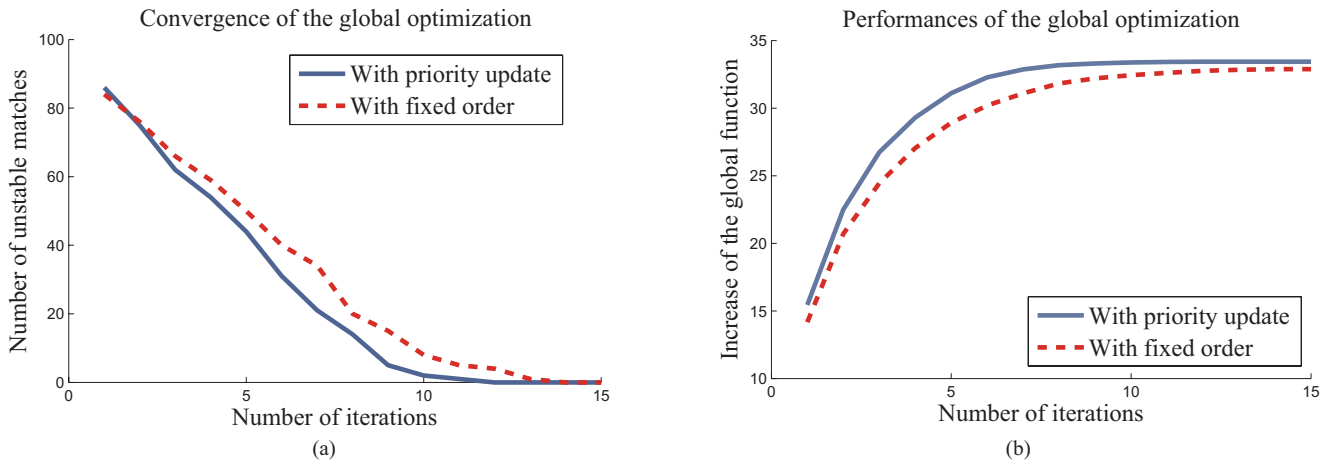


Fig. 6. Performances of global optimizations using different order strategies. (a) Number of unstable matches decreases through iterations. (b) Global function value increases through iterations.

Fig. 7. Image pairs used for the test of match refinement and propagation. (i): (a) *Bowel*, (b) *Face*. (ii): (a) *Yoga*, (b) *Girl*. (iii): (a) *Michelle*, (b) *Blonde*. (iv): (a) *Ovo*, (b) *Church*.

2) For each candidate match $(\mathbf{f}_R^k, \mathbf{f}_T^k) \in \Phi^n$, define an propagation attempt $(\mathbf{f}_R^n, \mathbf{f}_T^{nk})$ with $S_T^{nk} = \mathbf{A}\mathbf{f}^k S_R^n$. Then evaluate the quality of $(\mathbf{f}_R^n, \mathbf{f}_T^{nk})$ by $f(S_R^n, \mathbf{A}\mathbf{f}^k S_R^n))$, i.e. $f(\mathbf{A}\mathbf{f}^k)$ in (9), where the affinity set Ψ^k (8) is selected from the seed matches Θ only;
3) Find the best propagation attempt as the initialization of match $(\mathbf{f}_R^n, \mathbf{f}_T^n)$.

$$(\mathbf{f}_R^n, \mathbf{f}_T^n) = \arg\max_k f(\mathbf{f}_R^n, \mathbf{f}_T^{nk}) \tag{12}$$

2) Refinement:

1) Simultaneously Refine all new matches $(\mathbf{f}_R^n, \mathbf{f}_T^n)$ by maximizing the global function $F(\{\mathbf{A}\mathbf{f}^k\})$ (7);
2) Apply the affine consistency filter (Section III) to remove the outliers from the whole set of matches.

V. Experimental Results

A. Image Matching

In this section, we intend to demonstrate the efficiency of the proposed global match refinement and propagation method. To this end, we compare the image matching results obtained by the global method with those produced by the local method [5]. The image pairs used for this evaluation are shown in Fig. 7, each contains a reference image with a well defined object and a target image where the object is transformed or deformed and possibly with clutters. According to the degree of texture and the transform between the reference and target images, the eight image pairs are classified into four categories: (i) weakly textured+non-rigid deformation (*bowel*, *face*); (ii) weakly textured+rigid transformation (*yoga*, *girl*); (iii) highly textured+non-rigid deformation (*michelle*, *blonde*); (iv) highly textured+rigid transformation (*ovo*, *church*). Thus this experiment could also give us an indication how these two factors, texture and image transform, will affect the performance of match refinement and propagation.

To provide a fair comparison, both local and global methods start with the same initial matches of Harris and Hessian Affine features (refer to Section III-C for details). Besides, the search range in the affine space (t_x, t_y, s_x, s_y, θ, h) and the parameter r used to sample new features are set the same for both methods. The differences of the two lie in that the local method will first locally refine each initial match to produce a seed match, and then propagate the new features individually based on the seed matches, whereas the global method will apply the proposed global optimization to both the refinement of initial matches and the propagation of new features.

Two different measures are used to quantify the performance of the local and global approaches. The first one, called *ratio of correct matches*, measures the success rate of match propagation. It is computed by the number of the new features that are correctly matched to the target image with respect to the total number of new features sampled in the reference image (do not include those with empty candidate set Φ^n (11)). It is difficult to define the ground truth of feature matches especially for the case of non-rigid deformation.

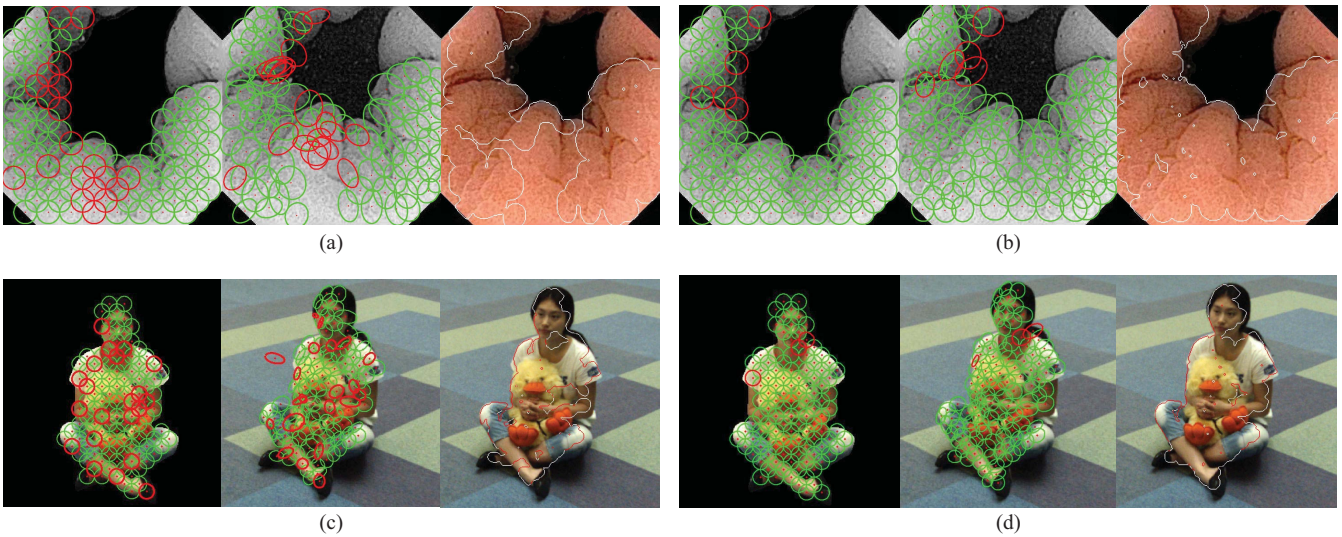


Fig. 8. Feature matching results by match propagation. (a) *Bowel* image pair, by local method. (b) *Bowel* image pair, by global method. (c) *Girl* image pair, by local method. (d) *Girl* image pair, by global method.

In this experiment, the correctness of a match is visually judged in terms of the consistency of its corresponding keypoints and support regions. To alleviate possible bias and errors in this judgement, we ask 10 persons to tell the correctness individually and only one match is shown at one time for the ease of observation. Then a match with at least 6 votes from the 10 persons are deemed as a correct one. Besides, for the rigid cases (ii) and (iv) an initial epipolar filter is applied to help identify the false matches. We first manually mark a few point matches as ground truth to estimate the fundamental matrix. Then a propagated match that violates the estimated epipolar geometry, i.e., one pixel deviated from the epipolar line, is automatically detected as a false match before the subjective judgement.

The second measure *region coverage* is introduced because we are not only concerned with the number of correct matches, but also interested in to what extent the object region in the target image is covered by the correctly propagated features. For this purpose, we first manually mark in the target image the ground truth region R_g that should have correspondence in the reference image. We then measure the region R_p which is composed by superimposing the support regions of all correctly propagated features. Thus *region coverage* is defined by (13), where $R_g \cap R_p$ and $R_g \cup R_p$ represent the intersection and union of the two regions, respectively. This measure will favor the case that R_p mostly conforms to R_g

$$region\ coverage = \frac{R_g \cap R_p}{R_g \cup R_p} \tag{13}$$

Experimental results in terms of *ratio of correct matches* and *region coverage* are summarized in Table II. The global method consistently outperforms the local method in terms of both measures, especially for the weakly textured image pairs, including *bowel*, *face*, *yoga* and *girl*. While for highly textured image pairs, i.e., *michelle*, *blonde*, *ovo* and *church*, the two have close performance. In average, the global method achieves nearly 10 percent better in terms of both measures.

TABLE II
PERFORMANCE OF MATCH REFINEMENT AND PROPAGATION

Image Pairs	$R_{\text{correctmatch}}$		R_{coverage}	
	Local	Global	Local	Global
Bowel	0.812	0.948	0.704	0.865
Face	0.747	0.892	0.659	0.819
Yoga	0.761	0.937	0.734	0.887
Girl	0.802	0.955	0.775	0.894
Michelle	0.908	0.936	0.816	0.882
Blonde	0.914	0.943	0.827	0.876
Ovo	0.971	0.979	0.890	0.901
Church	0.868	0.922	0.797	0.848
Average	0.847	**0.939**	0.775	**0.871**

As expected, the proposed affine consistency does help regularize match propagation especially when local appearance is less reliable. This can be observed as well in Fig. 8, where the matching results of *bowel* and *girl* image pairs are visually displayed. In Fig. 8 the green ellipses represent the correctly matched features and the red ellipses show the false ones. We can see that the features propagated by the global method have better accuracy of keypoints and support regions, and in general present more smoothly changing facets due to the affine consistency imposed. Another observation is that mismatches produced by the local method mostly come from low texture surfaces, while most mismatches yield by the global method (and some by the local method) are due to the fact that the features' support regions are two complex to be modeled by planar surfaces.[5] Fig. 8 also shows the region coverage results in *bowel* and *girl* target images. The area included by the white-red boundary indicates the intersection of regions $R_g \cap R_p$. It can be clearly observed that the features propagated by the global method are able to cover more object surfaces. Regarding the factor of image transform, no clear distinction in performance can be observed in Table II. Both

[5]Specifically, the regions may contain strong surface discontinuity or be partially occluded in the target image

methods work well for highly textured scenes either with rigid transformation or non-rigid deformation. Whereas our global method is apparently superior for low texture scenes due to the regularization by affine consistency.

B. Object Recognition

In the most related work [5], Ferrari *et al.* applied their contraction and expansion (referred to as local propagation in this paper) method to simultaneous object recognition and segmentation, and demonstrated its success for highly textured objects. However, they also mentioned that their method may not be suitable for objects with weak texture. In comparison, we have shown in Section V-A that our method is able to handle both highly and weakly textured objects, and overall achieves better performance in the image matching experiment. In this section, the capability of our method to deal with low-texture and non-rigid objects is further demonstrated by an experiment of leaf recognition.

The model and test images are selected from the database used in the Smithsonian project.[6] There are four classes A, B, C and D, each with a single model image and five test images, as shown in Fig. 9. The leaf samples in these images are typically non-rigid and weakly textured, and the four classes are very similar to each other in appearance. In this experiment, recognition is done by matching all pairs of model and test images (totally $4 \times 20 = 80$ pairs), and counting the amount of matches. For comparison, image matching is done by four different methods, namely "*original*," "*filter*," "*local*" and "*global*." "*original*" method just performs descriptor matching of Affine Harris and Hessian features as in Section III-C. "*filter*" will further employ the proposed geometric filter to remove the outliers by "*original*." "*local*" will then apply the local propagation [5] based on the matches after filtering, while "*global*" uses the proposed global propagation instead of the local method.

The resulting ROC (receiver operating characteristic) curves for the four classes are presented in Fig. 10, which depict the detection rate versus false-positive rate while varying the detection threshold, i.e. the number of matches. It can be observed in Fig. 10 that the "*global*" method consistently outperforms the other methods for all the four classes. In comparison, the performance of "*local*" method is quite unstable. In Fig. 10(a), its ROC is even significantly lower than "original" and "filter." We believe the main reason is that the local propagation generates a lot of false matches due to the poor discriminative power of leaf textures.

In Table III, we give the overall recognition rate. A test image is deemed to be successfully recognized if its matches found in the corresponding model image is more than those found in the other three model images. Results in Table III suggests that the proposed mismatch filter and global match propagation can help improve the recognition rate of the low-texture leaves, while the local propagation cannot help in this case.[7]

[6]Available at: http://www1.cs.columbia.edu/cvgc/efg/.

[7]Note that "*local*" is based on "*filter*" but its performance is worse.

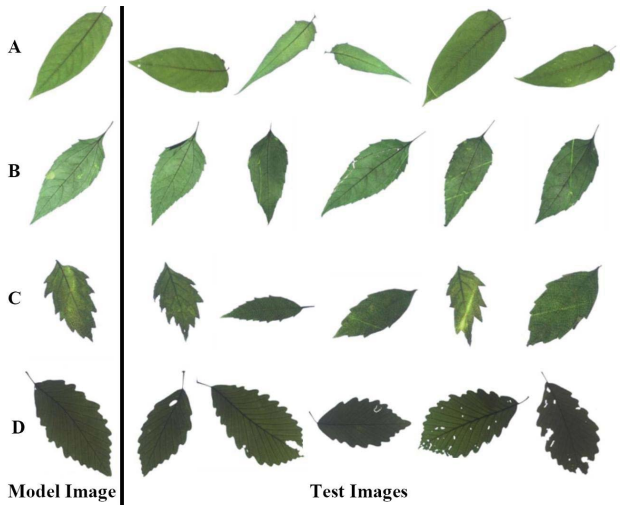


Fig. 9. Test and model images for leaf recognition.

TABLE III
RECOGNITION RATE BY DIFFERENT MATCHING METHODS

Original	Filter	Local	Global
0.60	0.75	0.65	0.85

C. Image Based Rendering

Lastly we extend the application to image based rendering. We show that due to the good accuracy and coverage of the features found by our method, a simple mesh modeling can be used to reconstruct and render the object surfaces from only two wide baseline images. Note that most existing automated 3D reconstruction approaches [2], [17], [21], [22] either assume narrow baseline configuration or require multiple input images (at least three).

Our feature-based modeling and rendering approach involves several steps. First, a sparse set of feature correspondences is generated by descriptor matching as in Section III-C. Next, a quasi-dense set of matches is produced by performing the proposed affine consistency filter and global match propagation which are described in Section III and Section IV, respectively. Note that in match propagation, epipolar constraint for rigid scene can be used to reduce the search space of 2D translation (t_x, t_y) to 1D disparity along the epipolar line. Based on the quasi-dense matches, a 3D triangular mesh model can then be constructed with the cameras fully calibrated. By taking into account the fact that the sampling rate of new features is still limited and the object surface may be very complex, we employ the image consistent surface triangulation [14] to find the best mesh surface in the sense that the appearances of the meshes are most consistent between two views. Finally, for each triangular mesh, we map the textures of the two input views onto the novel view by homography, and then blend the two warped textures for rendering.

The feature-based rendering approach is tested on real image datasets where the images are captured from significantly different viewpoints.

Girl (659×493)*:* Fig. 11(a) and (b) show the two input images where the object is weakly textured. The angular spacing between the two views is more than 30 degree. To help define the object of interest, we additionally provide

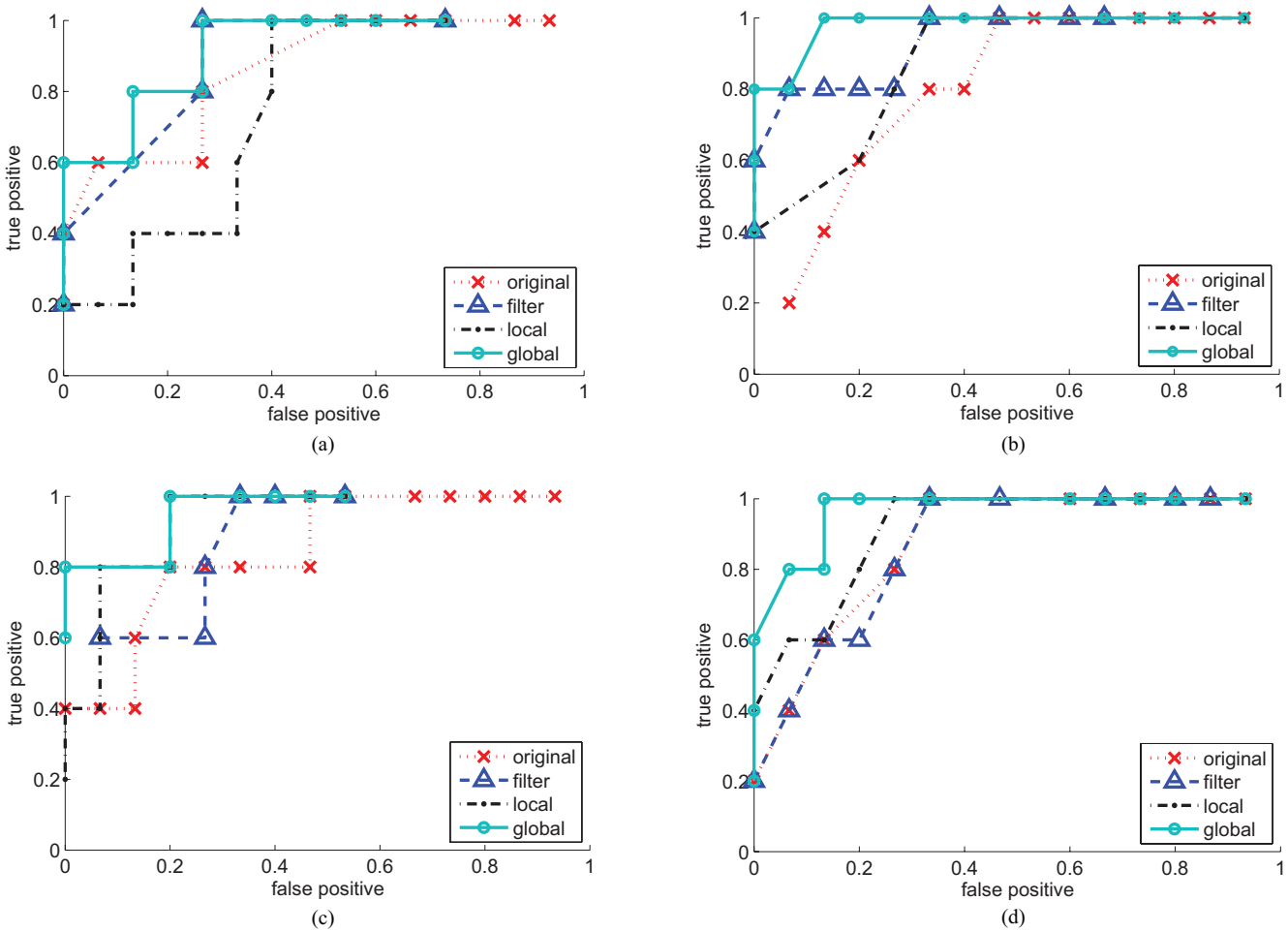


Fig. 10. ROC curves by different matching methods. (a)–(d) ROC for classes A, B, C, and D.

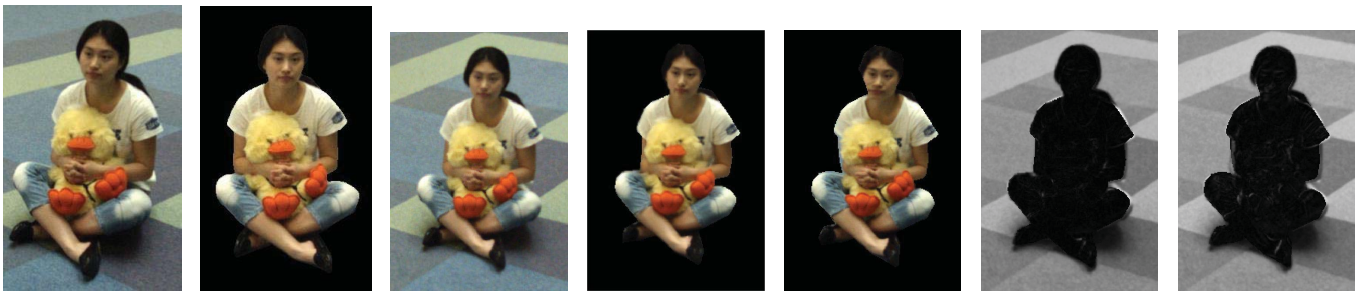


Fig. 11. Rendering results of *Girl sequence*. (a) and (b) Two input images. (c) Image of ground truth. (d) Virtual image of (c) by the proposed method. (e) Virtual image of (c) by local propagation. (f) Difference image between (c) and (d). (g) Difference image between (c) and (e).

a silhouette in the reference image as shown in Fig. 11(b). And in match propagation new features are also sampled along the silhouette to match the object boundary. The resulting look-around sequence of the novel views generated by the proposed method can be found at http://www.ee.cuhk.edu.hk/∼chcui/. Despite the difficulties of only two input images, widely separated views and weakly textured surfaces, the synthesized views are with high fidelity and the look-around sequence presents natural and smooth visual transition across views. In Fig. 11(c), an additional view is captured by the camera and used as the ground truth. Fig. 11(d) shows the virtual image of the object synthesized by our method, which is observed from the same viewpoint as in Fig. 11(c). The difference between the virtual image and the ground truth image is shown in Fig. 11(f), where the darker the image the smaller the magnitude of the RGB difference vector.[8] As we can see, the only noticeable differences are due to some self-occlusions that cannot be accurately modeled by the mesh facets. For comparison, Fig. 11(e) shows the rendering result by using the local propagation instead and Fig. 11(g) shows the difference image of (e) and the ground truth (c). In Fig. 11 (e) and (g) we can notice a few distortions around the girl's face, arms and fingers and the toy's feet, while Fig. 11(d) is free of such distortions.

[8]This measure is more sensitive than intensity difference.

(a) (b) (c) (d)

Fig. 12. Rendering results of *Cityhall* sequence. (a) and (b) Two input images. (c) and (d) Two virtual views synthesized by the proposed method.

Cityhall (768 × 512)*:* The original Cityhall sequence [18] consists of seven images of size 1536 × 1024. In this experiment, we choose only two of them as the input images shown in Fig. 12(a) and (b), and downsize the images to one fourth of the original size. Different from *Girl*, the *Cityhall* scene is well textured but the two input images present strong scale change and perspective deformation. This time the whole image of Fig. 12(a) is defined as the object of interest. Our method is then used to generate a video sequence that shows the visual transition from the viewpoint of Fig. 12(a) to that of Fig. 12(b). For the complete rendering sequence, please refer to http://www.ee.cuhk.edu.hk/∼chcui/. In Fig. 12(c) and (d), we gives two virtual views extracted from the sequence. It can be observed that most of the annoying distortions come from the occluded parts that inherently cannot be matched and well modeled. For the remaining parts, the visual quality is quite satisfactory. The texture details are well preserved and most of them are free of artifacts.

To conclude, we have shown in this experiment that modeling by a quasi-dense set of features found by our method is sufficient to generate realistic synthetic views. The sparsity of features can simplify the 3D representation and reduce the data storage and memory cost. The capability of handling few images with significant viewpoint change allows for flexible camera layout and can reduce the cost and effort of multi-view system setup.

D. Computational Complexity

The most time-consuming part of our method lies in the match propagation. Despite the number of initial matches, the convergence of match propagation generally takes 10 to 15 iterations throughout our experiments, and the curve shown in Fig. 6(a) is typical for most of our test images. Note that from Fig. 6(a) the number of unstable matches drops quickly, which means that the computations keep decreasing proportionally through iterations. It is reported that the local method [4] typically takes 3 to 10 iterations for each match. Let N be the number of matches to be refined and let us approximate the curve in Fig. 6(a) by a straight line. Thus the global method takes $5N$ to $7.5N$ evaluations of f (9)[9] over the bounded 6D space, while the local method takes $3N$ to $10N$ evaluations of the similarity over the same space. Since evaluating the similarity is much more expensive than evaluating the affine consistency, the global and local methods should have similar computational cost. This is verified by our experiment. Table IV summarizes the runtime of computing match propagation for the 80 image pairs in Section V-B. It is implemented in non-optimized MATLAB code and measured on a Core Duo T2400 1.83 GHz windows laptop. The result turns out to be that the global match propagation is faster than the local method [4] [5].

[9]A combined measure of similarity and affine consistency.

TABLE IV
RUNTIME (s) of Match Propagation

Method	Mean	Std	Max	Min
Global	298.6	86.3	611.9	150.0
Local	382.3	121.5	815.8	193.3

VI. CONCLUSION

Appearance-based matching is likely to fail for surfaces that are poorly textured. In this paper, we emphasize the combination of local appearance and local geometry to solve this problem. Traditional matching methods rely on appearance descriptors only, thus may produce a lot of mismatches when clutters and viewpoint changes become significant. To remove these outliers, we propose an efficient geometric filter based on the consistency of local affine geometry. The filter is demonstrated to have excellent performance even when the images contain extensive clutters and the objects undergo strong non-rigid deformation. We then propose a global optimization method to refine the remaining matches and to propagate more matches that can densely cover the object surface. By imposing the affine consistency of affinities in the global function, our method can successfully regularize the weakly textured regions and meanwhile respect the depth discontinuities. Experiments in image matching show that the proposed global method has close performance to the local method [5] for highly textured surfaces, and is superior in dealing with the low-texture surfaces. To demonstrate the applications, we apply the proposed geometric filter and global propagation method to object recognition and image based rendering. We show that the proposed method can help improve the recognition rate of non-rigid low-texture leaves, and is able to synthesize high quality novel views from only two images with significant viewpoint change.

REFERENCES

[1] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.
[2] D. Bradley, T. Boubekeur, and W. Heidrich, "Accurate multi-view reconstruction using robust binocular stereo and surface meshing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[3] G. Carneiro and A. D. Jepson, "Flexible spatial configuration of local image features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2089–2104, Dec. 2007.
[4] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Wide-baseline multiple-view correspondences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol 1. Jun. 2003, pp. 718–725.
[5] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation by image exploration," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 40–54.
[6] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[7] J. Kannala and S. S. Brandt, "Quasi-dense wide baseline matching using match propagation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[8] S. Lazebnik, C. Schmid, and J. Ponce, "Semi-local affine parts for object recognition," in *Proc. Brit. Mach. Vis. Conf.*, vol. 2. 2004, pp. 959–968.
[9] M. Lhuillier and L. Quan, "A quasi-dense approach to surface reconstruction from uncalibrated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 418–433, Mar. 2005.
[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[11] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, 2004.
[12] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[13] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1, pp. 43–72, 2005.
[14] D. D. Morris and T. Kanade, "Image-consistent surface triangulation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2000, pp. 332–338.
[15] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or 'How do I organize my holiday snaps?"' in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 414–431.
[16] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 5, pp. 530–535, May 1997.
[17] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2006, pp. 519–528.
[18] C. Strecha, T. Tuytelaars, and L. Van Gool, "Dense matching of multiple wide-baseline views," in *Proc. Int. Conf. Comput. Vis.*, vol 2. 2003, pp. 1194–1201.
[19] P. H. S. Torr and D. W. Murray, "The development and comparison of robust methods for estimating the fundamental matrix. *Int. J. Comput. Vis.*, vol. 24, no. 3, pp. 271–300, 1997.
[20] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 61–85, 2004.
[21] A. Zaharescu, E. Boyer, and R. Horaud. "Transformesh: A topology-adaptive mesh-based approach to surface evolution," in *Proc. 8th Asian Conf. Comput. Vis.*, Nov. 2007, pp. 166–175.
[22] C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *Int. J. Comput. Vis.*, vol. 75, no. 1, pp. 49–65, 2007.

Chunhui Cui received the B.E. and M.E. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from the Chinese University of Hong Kong, Hong Kong, in 2010.

His current research interests include image and video processing, image and video coding, and computer vision.

King Ngi Ngan (F'00) received the Ph.D. degree in electrical engineering from Loughborough University, Loughborough, U.K.

He is currently a Chair Professor with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong. He was a Full Professor with the Nanyang Technological University, Singapore, and the University of Western Australia, Perth, Australia. He was a Honorary Professor or a Visiting Professor with numerous universities in China, Australia, and South East Asia. He has authored or co-authored over 300 refereed papers in journals and conferences, authored three books, edited six volumes, and edited nine special issues in journals. He holds ten patents on image and video coding and communications.

Prof. Ngan was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *Journal on Visual Communications and Image Representation*, the *EURASIP Journal of Signal Processing: Image Communication*, and the *Journal of Applied Signal Processing*. He was the Chair or the Co-Chair of a number of prestigious international conferences on image and video processing, including the 2010 IEEE International Conference on Image Processing, and was on the advisory and technical committees of numerous professional organizations. He is a fellow of the Institution of Engineering and Technology (U.K.) and the Institution of Engineers Australia, and was an IEEE Distinguished Lecturer from 2006 to 2007.