

Feature Matching with Bounded Distortion

Yaron Lipman¹, Stav Yagev¹, Roi Poranne¹, David W. Jacobs² and Ronen Basri¹

¹ Weizmann Institute of Science

² University of Maryland

We consider the problem of finding a geometrically consistent set of point matches between two images. We assume that local descriptors have provided a set of candidate matches, which may include many outliers. We then seek the largest subset of these correspondences that can be aligned perfectly using a non-rigid deformation that exerts a bounded distortion. We formulate this as a constrained optimization problem and solve it using a constrained, iterative reweighted least-squares algorithm. In each iteration of this algorithm we solve a convex quadratic program obtaining a globally optimal match over a subset of the bounded distortion transformations. We further prove that a sequence of such iterations converges monotonically to a critical point of our objective function. We show experimentally that this algorithm produces excellent results on a number of test sets, in comparison to several state-of-the-art approaches.

Categories and Subject Descriptors:

Additional Key Words and Phrases: image matching, feature correspondence, bounded distortion

1. INTRODUCTION

Finding corresponding points between pairs of images is one of the most fundamental problems in graphics and vision. Correspondences can be used to transfer information between images or models. They also form the basis for many approaches to three-dimensional reconstruction, object recognition and retrieval. Identifying corresponding points is difficult because the shape of objects and their appearance can change across images or models, as we may compare objects in different poses or even different objects from the same class. In such cases both the appearance of corresponding points and their context can differ significantly.

Recent years have seen a surge of methods for matching point features in the graphics and vision literature, see reviews in [Heider et al. 2011; Tuytelaars and Mikolajczyk 2008]. These methods identify “interest points” in each image or model and associate descriptors with them based on their local intensity or shape patterns. These descriptors are used subsequently to produce a collection of candidate matches. However, alongside correct correspondences these methods typically produce a considerable number of incorrectly matched outliers. For example, Figure 1(middle) shows two sets of potential corresponding pairs of points as extracted by the SIFT algorithm [Lowe 2004]. These sets contain many incorrect pairs and only relatively few good ones. (Corresponding points are marked by disks of the same color and size. Their unorganized pattern in the right image is due to the large number of incorrect matches.) Our method (right figures) selects a large subset of points that are consistent with a global deformation of bounded distortion.

To make these correspondences useful, we must remove outliers. Fortunately, these outliers are frequently geometrically inconsistent with the correct matches. Judging geometric consistency requires us to model how a configuration of points can change from one image or surface to another. Most commonly, a low-dimensional, parameterized model of deformations is used for this purpose. A

seminal example is RANSAC [Fischler and Bolles 1981], which is often applied to similarity or affine distortions, or distortions that obey epipolar constraints. However, many real-life correspondence problems require a high-dimensional deformation model to account for articulation of parts, non-rigid deformations of shape, or variations of shape between similar objects. This precludes the use of RANSAC and generally makes it more difficult to enforce geometric consistency on correspondences. More recent methods, which we discuss in Section 2, seek correspondences that respect these high-dimensional deformations. These methods can suffer from two problems. They may not offer a model of distortion that effectively captures the available geometric information, and they may lead to complex optimization problems that are attacked heuristically, or using relaxation. We offer an alternate approach to modeling distortion and optimizing matching.

Our method models deformations as bounded in extent. This offers a good compromise; while some deformations are needed to handle real-world effects, allowing arbitrary deformations loses all geometric constraint on corresponding points. We then filter correspondences as follows. Given a set of candidate pairs of correspondences as input, $\{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$, we aim to find the maximal subset, $\{(\mathbf{p}_{i_\ell}, \mathbf{q}_{i_\ell})\}_{\ell=1}^n$, $n \leq N$, that can be aligned with a low-distortion, non-rigid deformation. That is, if we denote by \mathcal{F}_K the collection of all deformations with distortion lower than or equal to some acceptable K (for some measure of distortion), we aim at solving

$$\min_{\Phi \in \mathcal{F}_K} \sum_{i=1}^N \|\Phi(\mathbf{p}_i) - \mathbf{q}_i\|^0, \quad (1)$$

where we use the notation $\|\cdot\|^0$ to denote the mixed 2, 0-norm, that is, the ℓ_0 -norm applied to the Euclidean distance between $\Phi(\mathbf{p}_i)$ and \mathbf{q}_i , so $\|\Phi(\mathbf{p}_i) - \mathbf{q}_i\|^0 = 1$ if $\Phi(\mathbf{p}_i) \neq \mathbf{q}_i$ and $\|\Phi(\mathbf{p}_i) - \mathbf{q}_i\|^0 = 0$ otherwise.

We propose a novel algorithm to perform this minimization. Our approach consists of solving a sequence of convex problems, each of which provides a better approximation to the limit case expressed in (1). Since at each iteration we have a convex problem, we can find its global minimizer and are therefore less sensitive to local minima than classical gradient descent methods. To build the convex problems we use three main ingredients: a) we approximate the (non-smooth) ℓ_0 functional in Eq. (1) using a family of smooth functionals, converging to it point-wise; b) we construct convex spaces that approximate the space \mathcal{F}_K of K -bounded-distortion deformations of triangular meshes following [Lipman 2012], which will form the domain for the optimization; and c) we use a specially tailored Constrained Iterative Reweighted Least Squares (IRLS) algorithm. The algorithm solves a series of globally convex quadratic programs to approximate minima of the smoothed ℓ_0 functionals over the convex bounded-distortion deformation spaces. We further prove that our algorithm converges monotonically to a critical point of these smoothed functionals, and that it returns a deformation that is guaranteed to be bijective with a bounded conformal distortion.

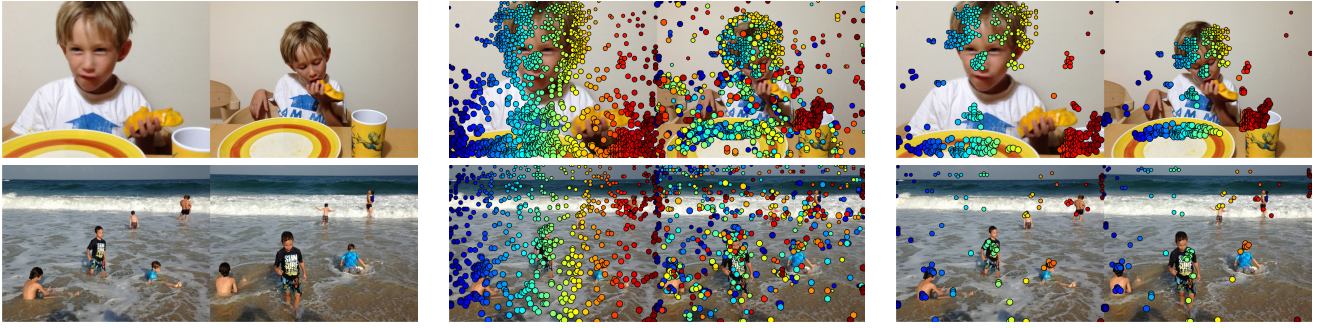


Fig. 1. A collection of candidate pairs of correspondences extracted with SIFT (middle column), and the subset of consistent pairs originated from a bounded-distortion deformation as produced by our algorithm (right column). Matching points are shown with disks of the same size and color.

We have tested our algorithm on different types of data, comparing it with several state-of-the-art algorithms for non-rigid matching and show superiority in precision and robustness to high rates of outliers.

2. PREVIOUS WORK

There has been a great deal of work on modeling and matching non-rigid objects [Montagnat et al. 2001; Nealen et al. 2006; van Kaick et al. 2011]. In this section we discuss work that specifically addresses the problem of matching point features in the presence of non-rigid deformations. We also discuss prior work on Iterative Reweighted Least Squares, which forms the basis for our algorithm for non-rigid matching.

2.1 Low dimensional deformation spaces.

In case the unknown deformation $\Phi(\mathbf{p}_{i_\ell}) \approx \mathbf{q}_{i_\ell}$, of a subset $\{(\mathbf{p}_{i_\ell}, \mathbf{q}_{i_\ell})\}_{\ell=1}^n$ is known to belong to some low dimensional deformation space \mathcal{D} (e.g., \mathcal{D} can be the group of affine maps, rigid motions, homographies, or deformations that are consistent with epipolar constraints), RANSAC-type [Fischler and Bolles 1981; Brown and Lowe 2003] methods can be used to robustly find this subset. In a nutshell, RANSAC aims at finding a large subset of pairs that are consistent with a deformation in \mathcal{D} up to an ε -deviation. Following the point of view and notations of Eq. (1), let \mathcal{D}_ε denote all deformations \mathcal{D} up to an arbitrary perturbation by ε , then RANSAC can be formulated precisely as (1), replacing the deformation space \mathcal{F}_K with \mathcal{D}_ε ,

$$\min_{\Phi \in \mathcal{D}_\varepsilon} \sum_{i=1}^N \|\Phi(\mathbf{p}_i) - \mathbf{q}_i\|^0. \quad (2)$$

\mathcal{D}_ε can also be viewed as a set of bounded deformations, but typically a small value of ε is used, producing a close approximation to a low-dimensional deformation space. Larger values of ε would lead to a large, unstructured deformation space. In that sense our method is a generalization of RANSAC to a set of low-distortion non-rigid deformations \mathcal{F}_K , which in sharp contrast to \mathcal{D} has (arbitrary) high dimensionality.

2.2 Minimizing deformations.

One way to deal with deformations that cannot be characterized by a small number of parameters is to use an energy function to measure the complexity of a deformation, and then seek correspondences that maximize the quality of the match while also minimizing the complexity of the deformation. [Chui and Rangarajan 2003]

use an iterative algorithm to minimize a deformation energy based on thin-plate splines, while also minimizing the distance between each point and the deformed point that matches it. [Belongie et al. 2002] optimize a similar energy, using shape context local descriptors to assist in matching points. [Jian et al. 2005] perform a related, non-rigid registration of point sets after first representing the point sets using mixtures of Gaussians. [Hinton et al. 1993] perform digit recognition using an elastic matching algorithm to minimize an energy that includes a spline-based deformation cost and a generative model of appearance. While these methods all allow for some outliers, they rely on algorithms that iterate between assigning correspondences and determining shape or deformation. This may make them sensitive to the quality of the initialization and to large numbers of outliers.

2.3 Graph-based methods.

Another approach is to model deformations based on local relationships between nearby pairs or triples of points. This approach particularly lends itself to graph-based matching approaches, in which pairwise relations between points can be represented by edges. For example, [Berg et al. 2005] optimize a cost that combines the similarity of geometric blur descriptors and the distortion in the geometric relationship between pairs of correspondences. A relaxation changes the formulation to a simpler, yet globally solvable, problem. [Lordecanu and Hebert 2005] form the weighted adjacency matrix of a graph that represents such costs, and use spectral relaxation, where the problem is posed as an eigenvector problem that can efficiently be solved globally. [Zheng and Doermann 2006] form a different graph, based on qualitative, neighborhood relationships, and use relaxation labeling to find correspondences that preserve these relationships. [Duchenne et al. 2011] form a hypergraph that encodes higher order geometric constraints between tuples and use an adjacency tensor to represent these relations. They also use a spectral relaxation to solve this. Another approach is to work directly with the low-distortion deformation space and perform combinatorial optimization [Gelfand et al. 2005; Funkhouser and Shilane 2006], or use greedy optimization [Tevs et al. 2009; Bronstein et al. 2006]. In both cases only a small fraction of the possible correspondences can be searched in a reasonable amount of time and the solution can be easily stuck in a local minimum. In general, graph-based formulations of matching lead to NP-hard problems that are attacked either with search or with a relaxation. Our approach, in contrast, represents the deformation globally as a bijective map, using piecewise affine transformations, while measuring the amount of distortion based on the relationships between triples of correspondences.

2.4 Patch-based methods.

Recently, patched based methods have been incorporated with a local geometric consistency prior to achieve dense correspondences between images. Patch-match [Barnes et al. 2009] looks for corresponding patches in the vicinity of a corresponding patch to improve geometric consistency. [HaCohen et al. 2011] generalize this to include color changes and other transformations beyond translation. Since patch-match generally grows corresponding regions in a bottom-up fashion it can easily produce globally inconsistent correspondences due to symmetries or repeating/similar structures in the image. Our method imposes global consistency, which can account for repeating patterns.

2.5 Iterative Reweighted Least Squares.

IRLS is a common method for minimizing ℓ_p norms, $p \geq 0$ by solving a sequence of convex quadratic problems [Lawson 1961]. Constrained IRLS has been used recently in the context of sparse coding and compressed sensing [Daubechies et al. 2010; Chartrand and Yin 2008], however in these cases the constraints are affine spaces and not more general convex sets. Some works generalize IRLS to the constrained setting [Eckhardt 1980; Bissantz et al. 2009] and provide some theoretical justification. We follow the general construction in these papers, but provide a particular derivation for the non-convex case of ℓ_p , $0 < p < 1$, for which we also prove monotonic convergence to a local critical point. Lastly, we note that alternative ℓ_p optimization techniques exist aside from IRLS [Bouaziz et al. 2013].

3. APPROACH

Given as input candidate corresponding pairs of points in the plane $(\mathbf{p}_i, \mathbf{q}_i) \in \mathbb{R}^2 \times \mathbb{R}^2$, $i = 1, \dots, N$, our goal is to extract a maximal subset $\{(\mathbf{p}_{i_\ell}, \mathbf{q}_{i_\ell})\}$, $\ell = 1, \dots, n \leq N$ that can be aligned with a low-distortion deformation Φ .

We consider deformations that are piecewise affine. This means that Φ is determined by a triangulation of the planar point set, $\{\mathbf{p}_i\}$, along with the target location of each point, which we denote $\tilde{\mathbf{p}}_i = \Phi(\mathbf{p}_i)$. We abuse notation and denote by Φ the matrix of unknowns, namely $\Phi = (\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2, \dots, \tilde{\mathbf{p}}_N) \in \mathbb{R}^{2 \times N}$. We use a Delaunay triangulation of $\{\mathbf{p}_i\}$, and denote the triangulation $\mathbf{T} = (\mathbf{V}, \mathbf{F})$, where $\mathbf{V} = \{\mathbf{p}_i\}$ is the vertex set, and $\mathbf{F} = \{f_j\}$ is the face set, indexed by oriented triplets of vertices $j = (i_1, i_2, i_3)$ that are assumed to be not co-linear. The map Φ is then extended linearly to each face, e.g., the face f_j is mapped by the affine transformation $\mathcal{A}_j(\mathbf{p}) = A_j \mathbf{p} + \mathbf{t}_j$, where $\mathbf{t}_j \in \mathbb{R}^{2 \times 1}$, $A_j \in \mathbb{R}^{2 \times 2}$ are set uniquely by requiring that $\mathcal{A}_j(\mathbf{p}_\ell) = \tilde{\mathbf{p}}_\ell$, $\ell = i_1, i_2, i_3$. (Note that we assume $\mathbf{p}_i, \tilde{\mathbf{p}}_i, \mathbf{q}_i \in \mathbb{R}^{2 \times 1}$, that is, all are column vectors.) These equations can be summarized in the following linear system:

$$(\tilde{\mathbf{p}}_{i_1}, \tilde{\mathbf{p}}_{i_2}, \tilde{\mathbf{p}}_{i_3}) = [A_j \ \mathbf{t}_j] \begin{pmatrix} \mathbf{p}_{i_1} & \mathbf{p}_{i_2} & \mathbf{p}_{i_3} \\ 1 & 1 & 1 \end{pmatrix}. \quad (3)$$

Note that given point correspondences ($\{\mathbf{p}_i\}$ and $\{\tilde{\mathbf{p}}_i\}$) A_j and \mathbf{t}_j can readily be recovered by inverting the right-most matrix, and that the relation between $\tilde{\mathbf{p}}_i$ and A_j, \mathbf{t}_j is linear in $\tilde{\mathbf{p}}_i$. Let us denote this linear relation by $A_j = A_j(\Phi)$.

Our goal is to align as many of the $\tilde{\mathbf{p}}_i$'s as we can with the \mathbf{q}_i 's. This is equivalent to minimizing the functional:

$$E(\Phi) = \sum_{i=1}^N \|\tilde{\mathbf{p}}_i - \mathbf{q}_i\|^0, \quad (4)$$

with the mixed $\ell_{2,0}$ -norm defined in (1). In general, we cannot match all pairs of points because we ban matchings that induce high distortion of our triangulation \mathbf{T} .

Next, we quantify the distortion of triangular meshes. Consider two triangular faces $\Delta = \Delta(\mathbf{p}_{i_1}, \mathbf{p}_{i_2}, \mathbf{p}_{i_3})$ and $\tilde{\Delta} = \Delta(\tilde{\mathbf{p}}_{i_1}, \tilde{\mathbf{p}}_{i_2}, \tilde{\mathbf{p}}_{i_3})$ related by the affine transformation $\tilde{\mathbf{p}} = \mathcal{A}_j(\mathbf{p}) = A_j \mathbf{p} + \mathbf{t}_j$. We measure the distortion induced by \mathcal{A}_j using the conformal distortion of its linear part $\mathbf{D}(A_j)$ defined as

$$\mathbf{D}(A_j) = \frac{\sigma_{\max}(A_j)}{\sigma_{\min}(A_j)}, \quad (5)$$

where $\sigma_{\max}(A_j)$ and $\sigma_{\min}(A_j)$ respectively are the maximal and minimal singular values of A_j . Clearly, $\mathbf{D}(A_j)$ is invariant to scale changes and measures deviation from similarity. In particular, $\mathbf{D}(A_j) = 1$ when (and only when) Δ and $\tilde{\Delta}$ are perfectly similar triangles, and otherwise $\mathbf{D}(A_j) > 1$. Note that this measure differs from [Duchenne et al. 2011]'s who compared triangles by the sine of their angles. We also do not want to allow any triangle to be reflected (flipped), leading to loss of injectivity of the deformation Φ . A triangle f_j is reflected if and only if $\det(A_j) < 0$.

Going back to Eq. (4) we now add these low-distortion constraints to complete our problem formulation:

$$\begin{aligned} \min_{\Phi} \quad & E(\Phi) & (6) \\ \text{s.t.} \quad & \begin{cases} \mathbf{D}(A_j(\Phi)) \leq K \\ \det(A_j(\Phi)) > 0 \\ \Phi \in \mathbb{R}^{2 \times N} \end{cases} & (7) \end{aligned}$$

where A_j is expressed as a constant linear combination of the unknowns $\tilde{\mathbf{p}}_i$, as can be computed from Eq. (3).

The optimization problem (6)-(7) presents several challenges: a) its functional E is not smooth and is non-convex; and b) the set of constraints (bounded conformal distortion, and no reflections) is non-convex. To solve this problem we introduce a hierarchy of simpler (smooth and convex) problems that converge locally to the non-convex problem above.

We separately treat the functional in Eq. (6) and the constraints in Eq. (7). First, the constraints (7) are replaced with a convex subset of constraints using a recent characterization of the space of bounded distortion mappings of triangular meshes [Lipman 2012]. Second, we approximate the ℓ_0 functional E in Eqs. (4),(6) by a sequence of smooth functionals E_δ and optimize each functional in this sequence using a generalization of the method of Iterative Reweighted Least Squares (IRLS), where a quadratic functional is solved over the convex space of bounded distortion deformations. Using these, we construct a straightforward algorithm and prove that it converges monotonically to a critical point of E_δ . Furthermore, the deformation returned by the algorithm is guaranteed to be bijective with a conformal distortion bound of K .

Next, we review the construction of the convex bounded distortion deformation spaces, and then present the generalized IRLS that furnishes our algorithm.

3.1 Convex bounded-distortion mapping spaces

In this section we review the space of bounded distortion mappings of triangular meshes in the plane as discussed in [Lipman 2012]. We reformulate and adapt the equations to our setting.

A 2×2 matrix A can be uniquely decomposed as follows

$$A = \frac{A - A^t + \text{tr}(A)I}{2} + \frac{A + A^t - \text{tr}(A)I}{2} = B + C, \quad (8)$$

where B is a similarity matrix (i.e., rotation and isotropic scale) and C is an anti-similarity matrix (i.e., a similarity matrix composed with a reflection).

The matrix A will have bounded conformal distortion of K and will not be a reflection (i.e., satisfies $\mathbf{D}(A) \leq K$ and $\det(A) \geq 0$) if its reflection component C is sufficiently small with respect to its similarity component B , or, more precisely, if

$$\|C\|_F \leq \frac{K-1}{K+1} \|B\|_F, \quad (9)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. In case we want strict inequality, i.e., $\det(A) > 0$, we should add to Eq. (9) the requirement that B (and hence A) is not the zero matrix. However, as we explain later, in our case either all matrices A_j are zero together or none is zero, and hence we will use Eq. (9) as is. Let us denote the set of all matrices A that satisfy Eq. (9) by \mathcal{C} .

Eq. (9) is non-convex due to the use of the Frobenius norm on the right hand side of the equation. We will therefore consider the maximal convex subsets of \mathcal{C} . (A convex set is maximal in \mathcal{C} if no other convex subset of \mathcal{C} contains it.) These subsets are characterized by the requirement that the rotation angle is bounded, where we define the rotation angle of a matrix A as the rotation angle induced by its similarity component B , and denote it by $\text{angle}B$.

The collection of maximal convex subsets $\mathcal{C}_\theta \subset \mathcal{C}$ is parameterized by a choice of a reference rotation angle θ for all choices of possible rotations, $-\pi < \theta \leq \pi$. Let the matrix R_θ denote rotation by θ , \mathcal{C}_θ is the set of matrices

$$\mathcal{C}_\theta = \left\{ C + BR_\theta \left| \|C\|_F \leq \frac{K-1}{K+1} \frac{\text{tr}(B)}{\sqrt{2}} \right. \right\}, \quad (10)$$

where $\text{tr}(B)$ denotes the trace of the matrix B . Intuitively, in general $\frac{1}{\sqrt{2}}\text{tr}(B) \leq \|B\|_F$, and therefore if $A \in \mathcal{C}_\theta$ then in particular A satisfies Eq. (9). The set $\mathcal{C}_\theta \subset \mathcal{C}$ balances between the distortion and rotation it allows, i.e., the set of allowed rotations in \mathcal{C}_θ shrinks with the amount of distortion applied. Formally, the maximal rotation angle allowed in \mathcal{C}_θ is defined by the following equation

$$|\theta_{\max} - \theta| \leq \cos^{-1} \left(\frac{K+1}{K-1} \frac{k-1}{k+1} \right), \quad (11)$$

where k is the conformal distortion of a matrix A in \mathcal{C}_θ . That is, for similarity, $k = 1$, and hence the maximal rotation angle of similarities contained in \mathcal{C}_θ is always in the range $(\theta - \pi/2, \theta + \pi/2)$ regardless of the value K , while for matrices with conformal distortion $k = K$ no rotation except θ is allowed. Intermediate rotations are allowed for $k \in (1, K)$ as can be read from Eq. (11).

Eq. (10) formulate convex cone constraints and a second order cone program (SOCP) solvers can be used to optimize a convex quadratic functional over this space. This renders SOCP solvers suitable for the framework presented in this paper. However, it is currently easier to work with linear inequalities rather than cone constraints and several suggestions for such further relaxations can be found in [Lipman 2012; Bommès et al. 2013]. Such relaxations can be achieved by using equivalence of norms, for example of the 0 and 2-norms; for 2×2 matrices, $\|C\|_\infty \leq \|C\|_F \leq 2\|C\|_\infty$, where $\|C\|_\infty := \max\{|(C)_{k\ell}|\}$. This leads to the following alternative definition of \mathcal{C}_θ (similarly to Eq. (10)):

$$\mathcal{C}_\theta = \left\{ C + BR_\theta \left| \|C\|_\infty \leq \frac{K-1}{K+1} \frac{\text{tr}(B)}{2\sqrt{2}} \right. \right\}. \quad (12)$$

We note that this space: a) is smaller than the one defined via Eq. (10); and b) similarly to Eq. (10) have a “nesting” property:

any matrix $A \in \mathcal{C}_\theta$, for some θ is also contained in the convex subset $\mathcal{C}_{\theta'}$ with $\theta' = \text{angle}B$. This property has a key role in proving the monotonicity of our algorithm. Note that this property does not hold with the infinity-norm relaxed spaces in [Lipman 2012].

An alternative relaxation is suggested in Bommès et al. [2013] using the 1-norm. Using the 1-norm and different rotations for the similarity and anti-similarity parts, i.e.,

$$\mathcal{C}_{\theta,\psi} = \left\{ CR_\psi + BR_\theta \left| \|C\|_1 \leq \frac{K-1}{K+1} \frac{\text{tr}(B)}{\sqrt{2}} \right. \right\}, \quad (13)$$

where $\|C\|_1 = \sum_{k\ell} |(C)_{k\ell}|$. One can build a different relaxation of Eq. (10) that also posses the nesting properties and has the advantage of being able to represent any matrix in \mathcal{C} . However, this space did not prove advantageous on the experiments in this paper and so we stick with the definition in Eq. (12).

The above discussion concentrated on one matrix A . For a bounded-distortion, piecewise-affine deformation of the triangulation \mathbf{T} , we will have an angle θ_j defining a convex space \mathcal{C}_{θ_j} per face $f_j \in \mathbf{F}$. By $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{|\mathbf{F}|})$ we will denote the vector of all angles, and denote the space \mathcal{C}_θ to be the set of all piecewise-affine deformations Φ such that their matrices satisfy $A_j \in \mathcal{C}_{\theta_j}$, that is

$$\mathcal{C}_\theta = \{ \Phi \in \mathbb{R}^{2 \times N} | A_j(\Phi) \in \mathcal{C}_{\theta_j}, \text{ for all } f_j \in \mathbf{F} \}. \quad (14)$$

Lastly, the space of bounded distortions \mathcal{C}_θ contains locally injective mappings in the sense that no triangle is reflected, but it does not forbid fold-overs (e.g., two points belonging to different triangles mapped to the same point). To guarantee no fold-overs it is enough to make sure the boundary of the mesh \mathbf{T} is mapped bijectively (see proof in [Lipman 2012]). We achieve this by adding to the vertex set \mathbf{V} new vertices $\{\mathbf{p}_{i'}\}$ placed on the scaled bounding box of $\{\mathbf{p}_i\}$ (we used 130% scale). We took approximately \sqrt{N} such points. The new vertices $\mathbf{p}_{i'}$ are mapped to some new locations $\tilde{\mathbf{p}}_{i'}$ under the one rule that they all are mapped via the same affine transformation. That is,

$$\tilde{\mathbf{p}}_{i'} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{p}_{i'} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}, \quad (15)$$

where a, b, c, d, t_1, t_2 are all new unknowns. We do not restrict the affinity a, b, c, d, t_1, t_2 in any way. Any $\Phi \in \mathcal{C}_\theta$ that is not the constant mapping (i.e., mapping all points to a single point) can be shown to be a bijection. Therefore, either all $A_j(\Phi)$ are zero, or none is zero.

3.2 Optimizing the ℓ_0 deformation functional

In this section we deal with the functional E in our optimization problem defined in Eq. (4). The two main challenges are the fact that E is neither smooth nor convex.

Our first task is to replace the functional E by a smooth functional, so we can construct a method to approximate its minima. We will consider the following standard family of smooth functionals

$$E_{p,\delta}(\Phi) = \sum_{i=1}^N \left(\|\tilde{\mathbf{p}}_i - \mathbf{q}_i\|^2 + \delta \right)^{\frac{p}{2}}. \quad (16)$$

Note that for every $p > 0$ and $\delta > 0$ these functionals are smooth, and if we take a sequence $\delta, p \rightarrow 0$ such that δ approaches zero much faster than p (e.g., $\delta = e^{-1/p^2}$) then $E_{p,\delta}(\Phi) \rightarrow E(\Phi)$ for every Φ (point-wise convergence). In this paper we will set p to be a small constant. (Throughout this paper we used $p = 0.001$.) Henceforth, let us denote $E_\delta = E_{p,\delta}$ and treat p as constant. In the

Algorithm 1: Low-distortion correspondence filtering.

Input:

Candidate correspondences $(\mathbf{p}_i, \mathbf{q}_i) \in \mathbb{R}^2 \times \mathbb{R}^2, i = 1, \dots, N$
Distortion bound $K \geq 1$

Output:

Subset of pairs $(\mathbf{p}_{i_\ell}, \mathbf{q}_{i_\ell}), \ell = 1, \dots, n \leq N$
Bounded distortion bijective map Φ

Triangulate $\{\mathbf{p}_i\}_{i=1}^N$ to get $\mathbf{T} = (\mathbf{V}, \mathbf{F})$;

Set $\delta^{(0)} \approx \text{diam} \{\mathbf{p}_i\}$;

Set $\tilde{\mathbf{p}}_i = \mathbf{p}_i$;

Set weights $w_i^{(0)}$ according to Eq. (20);

Set $\theta_j^{(0)} = 0$;

$m = 0$;

while $\delta^{(m)} > \delta_{\min}$ and $m < m_{\max}$ **do**

while $E_{\delta^{(m)}}(\Phi^{(m)}) - E_{\delta^{(m-1)}}(\Phi^{(m-1)}) > \varepsilon$ **do**

$m = m + 1$;

 Set $\delta^{(m)} = \delta^{(m-1)}$;

 Set weights $w_i^{(m-1)}$ according to Eq. (20);

 Set angles $\theta_j^{(m-1)}$ according to Eq. (21);

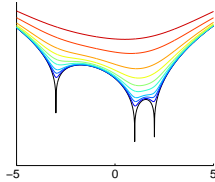
 Use quadratic programming to solve for $\Phi^{(m)}$ via Eqs. (17), (18);

 Set $\delta^{(m)} = \delta^{(m)}/2$;

Return Φ , and all pairs $(\mathbf{p}_i, \mathbf{q}_i)$ for which $w_i^\infty > \frac{1}{2}$;

evaluation section we show that changing the value p during optimization did not produce any advantage over keeping p constant.

Another difficulty in the functional E is that it is not convex. An important property of the functional family E_δ that helps in avoiding local minima is that its derivatives are inversely proportional to δ ; for large δ the derivatives of E_δ will be bounded by a small number (in a bounded domain), intuitively smoothing out shallow local minima. To visualize this effect, we show in the inset graphs of the functional $E_\delta(x) = \sum_{i=1}^3 (|x - q_i|^2 + \delta)^{\frac{p}{2}}$, where $x \in [-5, 5]$, $q_1 = -3, q_2 = 1, q_3 = 2$, and for different values of δ (top corresponds to large $\delta = 10$, bottom to small $\delta = 0.02$ values). In this example we took $p = 0.25$. The global minimum of this functional is achieved at $x = q_2 = 1$ and, as evident from the landscape of the reddish graphs, starting the optimization at large values of δ might help in avoiding the local minima at $x = -3, 2$.



Our algorithm starts with some initial large δ (e.g., proportional to the diameter of the image) and approximates the minimizer of E_δ subject to the constraints that the deformation Φ has bounded distortion K . Once we converge, we decrease δ by $\delta \leftarrow \delta/2$ and repeat the process, until δ becomes less than a threshold $\delta < \delta_{\min}$. Then we end up with a minimizer Φ of $E_\delta = E_{p, \delta_{\min}}$ with small δ_{\min} and p . Since this $E_{p, \delta_{\min}}$ approximates E point-wise this minimizer Φ is the desired output. The algorithm is summarized in Algorithm 1.

To optimize the functional E_δ with the bounded distortion constraints we use a generalization of Iterative Reweighted Least Squares (IRLS), solving a sequence of linearly constrained convex

quadratic problems:

$$\Phi^{(m)} = \underset{\Phi}{\operatorname{argmin}} E_\delta^{(m)}(\Phi) \quad (17)$$

$$\text{s.t. } \Phi \in \mathcal{C}_{\theta^{(m-1)}} \quad (18)$$

where,

$$E_\delta^{(m)}(\Phi) = \sum_{i=1}^N w_i^{(m-1)} \|\tilde{\mathbf{p}}_i - \mathbf{q}_i\|^2, \quad (19)$$

$$w_i^{(m-1)} = \left(\|\tilde{\mathbf{p}}_i^{(m-1)} - \mathbf{q}_i\|^2 + \delta \right)^{\frac{p}{2} - 1}, \quad (20)$$

$$\theta_j^{(m-1)} = \text{angle } B_j^{(m-1)}, \quad (21)$$

and $\mathcal{C}_{\theta^{(m-1)}}$ is defined in (14) (with the use of equations (12) and (15)). The matrix $B_j^{(m)}$ is the similarity component of $A_j^{(m)}$ (see Eq. (8)), and as before $\text{angle } B_j^{(m)}$ denotes its rotation angle. Intuitively, this angle sets the maximal convex space around the orientation determined by the solution of the previous iteration, allowing maximal flexibility in the deformation while staying in the convex regime.

For notational convenience we denote by $\delta^{(m)}$ the value of δ used in the m 'th iteration of the algorithm. Note that δ usually remains constant between iterations, except whenever $E_\delta(\Phi)$ converges. The algorithm can be shown to have desirable convergence properties and guarantees summarized in the following theorem,

THEOREM 1. *Algorithm 1 has the following properties. The energy $E(\Phi^{(m)}, \delta^{(m)}) \equiv E_{\delta^{(m)}}(\Phi^{(m)})$ is monotonically decreasing, that is*

$$E(\Phi^{(m+1)}, \delta^{(m+1)}) \leq E(\Phi^{(m)}, \delta^{(m)}),$$

and the series $E(\Phi^{(m)}, \delta^{(m)})$ converges to a critical value of the functional E_δ , for some $\delta \leq \delta_{\min}$ on a convex subset of the K -bounded distortion deformation space. Moreover, the algorithm is guaranteed to output either a constant solution or a bijective deformation Φ with K -bounded conformal distortion.

In fact it is readily possible to avoid the constant solution by adding another inequality in Eq. (10), but in practice we never encountered such a solution and it is unlikely to be a minimum in case the correspondences $\mathbf{p}_i, \mathbf{q}_i$ are general.

The proof of Theorem 1 is given in the Appendix. Current theory analyzing IRLS does not seem to encapsulate our case where both $0 < p < 1$ and the convex constraints are changed during iterations. To our knowledge, the closest theory deals with a convex functional and a fixed convex constraint space [Bissantz et al. 2009]. Therefore we generalize this analysis to include our scenario.

4. EXPERIMENTS AND EVALUATION

4.1 Comparisons to other techniques

We tested our method both in simulations and in experiments with real images and compared our method to a collection of techniques. Our Matlab implementation uses Matlab's quadratic programming routine to solve for $\Phi^{(m)}$ via Eqs. (17), (18). Throughout our experiments we use an ℓ_p norm with constant $p = 0.001$, and a distortion bound of $K = 3$. We further demonstrate below that our method is not sensitive to these choices. A final optional implementation comment is that once the algorithm converged, some points $\tilde{\mathbf{p}}_i$ may reach a location very close to their target \mathbf{q}_i but not to within the

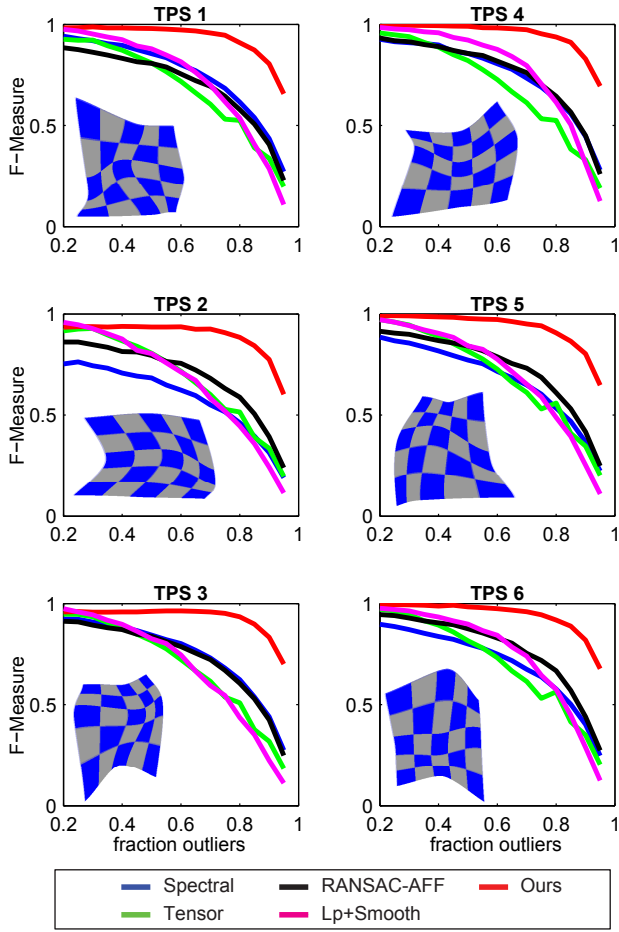


Fig. 2. Simulations with 6 out of the 24 TPS deformations used.

limits of numerical precision. This can happen when the particular triangulation prevents it. In these cases it is possible to snap these vertices in postprocessing (in our implementation we use the distance threshold of 5 pixels, where image typical size is 300×400 pixels).

Spectral-type. In this category we compared to the techniques of [Leordeanu and Hebert 2005] (denoted Spectral) and [Duchenne et al. 2011] (denoted Tensor), who use spectral relaxation techniques to minimize energies measuring consistency of relations between (respectively) pairs and triplets of points. The default design of each method allows any input point set to match any other one, returning a match for every point. We introduced slight modifications to enable the algorithms to select from candidate pairs of matches from the input and to reject outliers. Briefly, the cost matrix and tensor for the respective methods were populated using only the input candidate pairs (as opposed to all pairs in original methods). For the Spectral method, the cost function penalizes changes in distances between pairs of feature points. The cost function is minimized by solving an eigenvector problem. Then, to produce the final pairs of correspondences, we set a rejection threshold to their greedy post-processing step. Similarly, the Tensor method uses a cost function that penalizes changes in angles between triplets of feature points. Each triplet is represented by a 3-dimensional vector containing the sines of the three angles pro-

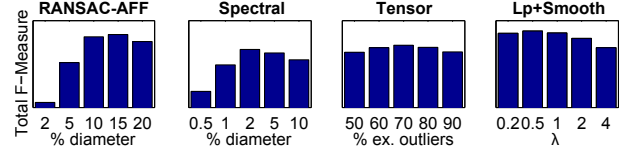


Fig. 3. Parameter choices for RANSAC-AFF, Spectral, Tensor and Lp+Smooth. The graphs show the performance (sum of area under the F-measure curve) over all 24 experiments as a function of parameter values. From left to right: (1) threshold for inlier distance for RANSAC-AFF as a fraction of the diameter, (2) threshold for postprocessing rejection for Spectral as a fraction of the diameter, (3) cost of matching of dummy node tuned to match a given fraction of outliers in Tensor, (4) weight of the smoothness term in the Baseline algorithm. The results shown for the synthetic experiments were those that gave maximum performance (15% for RANSAC-AFF, 2% for Spectral, 70% for Tensor, 0.5 for Baseline). The results shown on real images were obtained with identical parameters.

duced by the triplet and the cost function then penalizes differences between 2 such vectors. A spectral approximation is used to minimize this functional. Finally, we identify outlier matches by matching points to a dummy node (as the authors suggest).

RANSAC. In this category we compared to the RANSAC [Fischler and Bolles 1981] algorithm using an affine (denoted RANSAC-AFF) and epipolar (denoted RANSAC-EPI) model (using Matlab’s GeometricTransformEstimator object and estimateFundamentalMatrix function, respectively), allowing for a suitable perturbation threshold to handle deformations.

Lp+Smooth. To illustrate the need for having the bounded distortion constraints (Eq. (18)) instead of a more standard regularization term, we have implemented a version of our algorithm in which we dropped the BD constraints (18) and added instead the standard smoothness term $\lambda \int \|\Delta\Phi\|^2$, where Δ denotes the Laplacian operator. Specifically, let M denote the (diagonal) mass matrix and L the cotangent Laplacian of the triangulation \mathbf{T} , the smoothness term discretizes as $\lambda \text{tr}(\Phi^T L M^{-1} L \Phi)$ (see, e.g., [Jacobson et al. 2011; Lipman 2012]). For this algorithm λ serves as the main parameter balancing the matching energy and the map regularity. Similarly to our algorithm, we constrained the boundary points to undergo the same affine transformation.

Image based. This category includes algorithms that directly use image intensities to find correspondences. Here we used Large Displacement Optical Flow (LDOF) [Brox and Malik 2011], and patch-match [HaCohen et al. 2011]. As these algorithms use image intensities they do not address our problem exactly, but it is nevertheless instructive to compare to these popular techniques.

In all cases excluding the image based methods we have chosen the main parameters of each algorithm to maximize performance on a synthetic data set as described below, see Figure 3. We then used those parameters for the experiments with real images.

4.2 Synthetic Data

We tested our algorithm on synthetic non-rigid deformations that we produced as follows. We have produced 24 random Thin-Plate Splines (TPS) mappings of a square such that the mean aspect-ratio distortion plus twice its standard deviation is smaller or equal to 3. For each TPS we tested our algorithm by producing $n = 49$ inlier pairs of points by sampling uniformly from the square $\mathbf{p}_1, \dots, \mathbf{p}_n$ (as a jittered 7×7 grid) and applying the random TPS Ψ to this set of points, that is $\mathbf{q}_i = \Psi(\mathbf{p}_i)$. To produce a realistic outlier

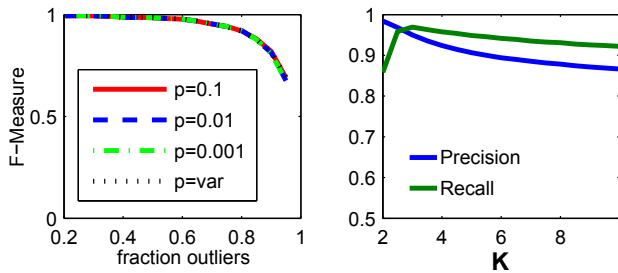


Fig. 4. Left: our method with different choices of quasi-norms p , and also for a version of the algorithm in which p is decreased during the optimization (graph shown with TPS-6). Right: The effect of varying the parameter K (graph shows the average performance obtained for TPS-6 with the different outlier fractions).

model, we added $N - n$ random outlier pairs $(\mathbf{p}_i, \mathbf{q}_i)$, $i = n + 1, \dots, N$ according to a distribution of outliers estimated from real images. Specifically, we constructed an error histogram for SIFT pairs $(\mathbf{p}'_i, \mathbf{q}'_i)$ in real images by measuring the distance of \mathbf{q}'_i to the ground-truth correspondence of \mathbf{p}'_i marked by a human observer. We then selected outlier pairs $(\mathbf{p}_i, \mathbf{q}_i)$ by first choosing \mathbf{p}_i , $i = n, \dots, N$ uniformly in the square, and then selecting \mathbf{q}_i randomly such that its distance to $\Psi(\mathbf{p}_i)$ follows the same distribution as the SIFT error histogram.

For each TPS we have created 100 random trials with an outlier fraction in the range $[0.2, 0.95]$. We scored every trial by its recall and precision rates and computed the F-measure as the harmonic mean of the average recall and precision rates over all trials as a function of fraction of outliers. Figure 2 shows typical 6 out of the 24 the TPS mappings along with a plot of F-measures obtained for our method compared to Spectral, Tensor, Lp+Smooth, and RANSAC-AFF. Our method outperforms the other methods in almost all cases. Moreover, the performance of our method remains high even with a very large fraction of outliers.

Sensitivity to parameters. We further tested the sensitivity of our algorithm to different parameter settings. In particular, we applied our method with several different choices of quasi-norm p in the $(0, 1)$ -interval as well as in a setup in which p is decreased during the optimization. Figure 4 shows that all of these choices lead to nearly identical performance. In addition we show the effect of varying K (the distortion bound) while holding the other parameters fixed at their chosen value. Clearly the selected values for K (i.e., $K = 3$) is sufficient to accommodate the tested deformation, but as the graphs show, the performance remains stable for a wide range of values.

4.3 Real images

Next we show results on real image pairs. We tested our method on a variety of image pairs depicting a variety of sceneries in three categories, images of dynamic scenes, images of static scenes, and images of different animals of the same breed. In all cases we found initial correspondences by matching SIFT descriptors (using the VLFeat software package [Vedaldi and Fulkerson]) and made sure that no point belongs to more than one SIFT pair. We then applied our method along with Spectral, Tensor, Lp+Smooth, LDOF, RANSAC-AFF and RANSAC-EPI to remove outliers. In all experiments we used the parameters that produced the best performance for the synthetic data (see Section 4.2, and Figure 3).

Our results are shown in Figures 5-7. To enhance visualization of matches we show matching points with disks of the same size and



Fig. 8. A series of pairs of frames with increasing deformation (top to bottom). The number of inliers detected with our algorithm decreases as the deformation grows larger.

color, with the color and size respectively varying according to the horizontal and vertical position of points in the left image. To further evaluate our results we had a human observer marking ground truth correspondences for all real images used. We show correspondence errors in each right image by connecting with a red line each correspondence chosen and the respective ground truth correspondence (to reduce clutter we do not use the red line markings in the SIFT images). The goal for each method therefore is to select as many of the SIFT pairs that are compatible with the ground truth or have small positional errors (indicated by short red lines) while discarding the ones that have large positional errors (indicated by long red lines). We further highlight with black disks good correspondences missed by the particular method tested. It is clear that our method outperforms the other methods in terms of both precision and recall.

Figure 5 (top) shows for example a scene where two subjects are moving away from each other. This is a challenging case as distances between feature points within each subject are preserved but distances across the subjects are not. Also the location of the subject w.r.t. the background is changing. Most methods miss one of the subjects (or alternatively accept many outliers) while our method is able to capture the full motion. In other examples in this Figure and Figures 6,7 please note that previous work either produce a high precision set of pairs but misses important chunks of good pairs (as indicated by the black disks in these figures), or achieves high recall rate (e.g., RANSAC-AFF) at the price of introducing quite a few false positives (see the long red lines to indicate such pairs). Figure 6 demonstrates the performance of our algorithm on static scenes that are captured from different viewpoints. These pairs are related geometrically by a fundamental matrix,



Fig. 9. Point matches obtained with the NRDC algorithm [HaCohen et al. 2011] (points are sampled uniformly from dense results). Notice sets of points incorrectly matched due to symmetry and repetitions (top), and unmatched areas due to large deformations (bottom).

and so corresponding points should lie along epipolar constraints. RANSAC-EPI finds the best fundamental matrix that fits the input pairs while discarding outliers. As the figure shows, it does so effectively, although it accepts some outlier matches when those happen to fall accidentally along their respective epipolar lines. Our algorithm, in contrast, is not constructed with static scenes in mind. Yet, as the figure shows, it performs reasonably well even in this case.

Figure 8 demonstrates the performance of our algorithm over a series of pairs of images that exhibit increasing deformation levels. As expected our method’s performance deteriorates as the deformation increases. This can be attributed to two things: first, SIFT tends to produce fewer inliers for larger deformations, and second, large deformations could violate our bounded distortion assumption.

We next relate our results to patch-based methods. Recent work in this directions aims to extract “local sets of dense correspondences between two images with some shared content” [HaCohen et al. 2011] by matching intensity patches (up to local deformations) with a local geometric consistency prior using the Patch-Match algorithm [Barnes et al. 2009]. This approach indeed often finds accurate matches, but, as we demonstrate in Fig. 9, can fail to correctly locate repeated structures or breaks in the presence of a strong deformation. We believe it would be interesting to complement these methods with geometric consistency as is done with our method.

Figure 10 shows the mesh \mathbf{T} (top) and its deformed version (bottom), as-well as highlights the subset of pairs selected by our algorithm (i.e. vertices that converged to the SIFT input pairs) by red disks.

Lastly, we present a table recording the precision, recall and F-measure rates achieved by the tested algorithms. We define the ground-truth set to include all SIFT corresponding pairs that deviate up-to 5 pixels from the marked ground-truth map. It is important

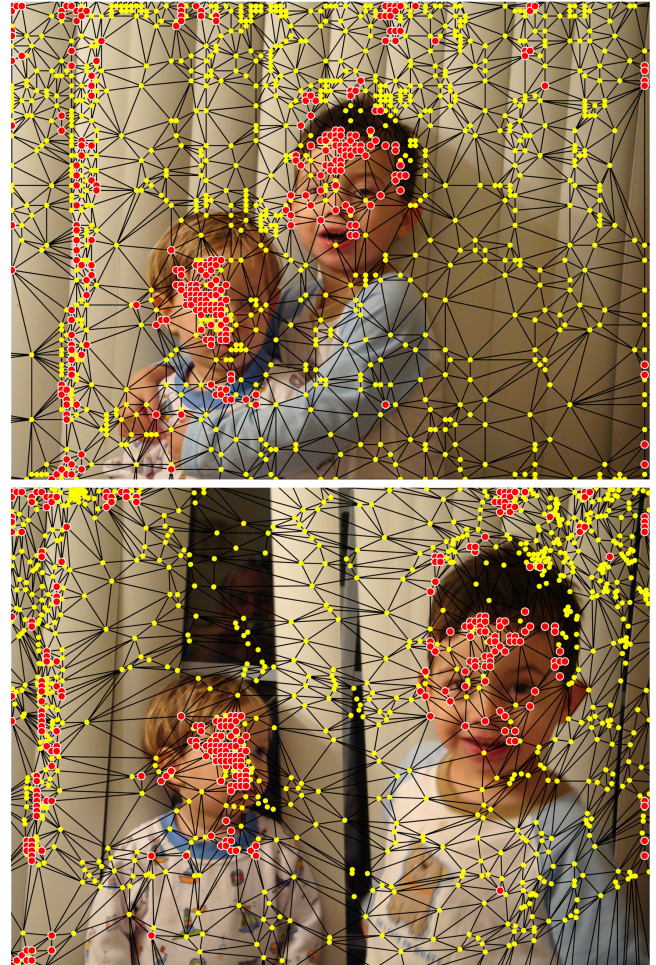


Fig. 10. The mesh before (top) and after (bottom) the deformation. The red disks depict “snapped” vertices (i.e., coincide with SIFT pairs as chosen by the ℓ_0 optimization).

to note that precision rates are overall relatively low for three reasons: 1) our ground-truth is defined according to a human observer and might have some discrepancies with some of the geometrically consistent SIFT pairs; 2) we set our pair’s acceptance threshold to be 5 pixels which is rather conservative; and 3) for the RANSAC methods, we took relatively large inlier acceptance parameter in order to be able to deal with larger distortions (see Figure 3).

method	Precision	Recall	F-measure
BD	56.0	94.0	70.1
L_p +Smooth	32.9	64.1	43.5
LDOF	49.5	20.2	28.7
RANSAC-AFF	41.9	99.7	59.0
RANSAC-EPI	42.5	54.9	47.9
Spectral	50.7	57.5	53.9
Tensor	36.9	27.7	31.6

Timings. Our algorithm requires solving a quadratic program in each iteration of the IRLS. Typical run times on Intel Xeon 2.3GHz are: 63s for $|\mathbf{V}| = 363$, 102s for $|\mathbf{V}| = 411$, 200s for $|\mathbf{V}| = 1043$.

5. CONCLUSION

This paper focuses on the problem of using geometric consistency as a filter, to separate outliers from correct matches between point features in images. We account for deformations in the apparent shape of objects by allowing for bounded conformal distortion in the configuration of the points. From a modeling perspective, bounded distortion allows us to account for some deformation while still retaining much of the geometric information that enables us to find correct correspondences. Algorithmically, bounded distortion is attractive because it enables us to approximate the set of distortions with maximal convex sets. Using this approximation, an iterative, reweighted least squares optimization can find good solutions. This is because our algorithm proceeds by finding the globally optimal solution to more and more accurate approximations to the complete, non-convex problem of aligning a maximal set of points with limited distortion. We are also able to prove that our algorithm converges to a critical point of this non-convex problem. We further show experimentally that our algorithm significantly outperforms several state-of-the-art algorithms.

We feel that our algorithm makes effective use of geometric constraints in removing outliers in matching. However, matching images in which there are significant variations due to changes in shape or other viewing conditions remains a challenging problem with many interesting open questions. For example, when deformations and lighting changes are large, existing descriptors such as SIFT may not be stable enough to propose a sufficient number of correct matches. Building descriptors that are also resilient to bounded conformal distortion would be an interesting direction for future work. Another limitation of our algorithm is that in scenes where the deformation is discontinuous our model can only supply sub-optimal results because it can only represent continuous deformations. The question of how to generalize to non-continuous deformations is also left for future work.

APPENDIX

THEOREM 1. *Algorithm 1 has the following properties. The energy $E(\Phi^{(m)}, \delta^{(m)}) \equiv E_{\delta^{(m)}}(\Phi^{(m)})$ is monotonically decreasing, that is*

$$E(\Phi^{(m+1)}, \delta^{(m+1)}) \leq E(\Phi^{(m)}, \delta^{(m)}),$$

and the series $E(\Phi^{(m)}, \delta^{(m)})$ converges to a critical value of the functional E_{δ} , for some $\delta \leq \delta_{\min}$ on a convex subset of the K -bounded distortion deformation space. Moreover, the algorithm is guaranteed to output either a constant solution or a bijective deformation Φ with K -bounded conformal distortion.

To prove this theorem we need the following theorem, which shows that the inner loop of Algorithm 1, in which δ is fixed, converges:

THEOREM 2. *The iteration described by (17),(18) produces monotonically decreasing functional values $E_{\delta}(\Phi^{(m+1)}) \leq E_{\delta}(\Phi^{(m)})$, and the sequence of values $E_{\delta}(\Phi^{(m)})$ converges to a critical value of the functional E_{δ} on a convex subset of the K -bounded distortion deformation space.*

Given Theorem 2 we can readily prove Theorem 1:

PROOF. (Theorem 1) Let us denote by $e_{\delta}^m = E(\Phi^{(m)}, \delta^{(m)}) = E_{\delta^{(m)}}(\Phi^{(m)})$ the series produced by the algorithm. For a fixed $\delta^{(m)} = \delta$, Theorem 2 implies that $e_{\delta}^m \geq e_{\delta}^{m+1} \geq 0$ and therefore

has to converge. At convergence (that is, for sufficiently large m) the algorithm sets $\delta^{(m)} := \delta^{(m)}/2$ and continues. By the definition of E_{δ} we have

$$e_{\delta}^m \geq e_{\delta/2}^m.$$

Therefore, the series e_{δ}^m is monotonically decreasing and not smaller than zero throughout the execution of the algorithm. Therefore, it converges. When the algorithm terminates the entire series e_{δ}^m converged to a critical value of E_{δ} for some $\delta \leq \delta_{\min}$ (by Theorem 2).

By construction, the final returned $\Phi^{(m)} \in \mathcal{C}_{\theta^{(m-1)}}$ and so in particular has conformal distortion $\leq K$. Also since it includes no reflected triangles and since the boundary is mapped bijectively (when excluding the case that the boundary is mapped to a single point, leading to the constant solution) it can be shown to be a bijection. \square

We now prove Theorem 2.

PROOF. (Theorem 2) The proof consists of several parts: (i) we show the monotonicity of the (non-convex) energy $E_{\delta}(\Phi^{(m+1)}) \leq E_{\delta}(\Phi^{(m)})$; (ii) we show that $\Phi^{(m)}$ has a convergent subsequence $\Phi^{(m_k)} \rightarrow \Phi^*$; (iii) we show that $\Phi^* \in \mathcal{C}_{\theta^*}$ for some angles θ^* ; and (iv) we show that Φ^* is a critical points of E_{δ} over the convex space \mathcal{C}_{θ^*} .

Part (i): Let us define the following functions similarly to [Bisantz et al. 2009]

$$h(r) = (r^2 + \delta)^{\frac{p}{2}} \quad (22)$$

$$E_{\delta}(\Phi) = \sum_{i=1}^N h(\|\tilde{\mathbf{p}}_i - \mathbf{q}_i\|) \quad (23)$$

$$g(r, s) = h(r) + \frac{p}{2} (r^2 + \delta)^{\frac{p}{2}-1} (s^2 - r^2) \quad (24)$$

$$G(\Phi, \Phi') = \sum_{i=1}^N g(\|\tilde{\mathbf{p}}_i - \mathbf{q}_i\|, \|\tilde{\mathbf{p}}'_i - \mathbf{q}_i\|), \quad (25)$$

where $\Phi = (\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2, \dots, \tilde{\mathbf{p}}_N) \in \mathbb{R}^{2 \times N}$ are unknown (and Φ' is defined similarly). Note that

$$\Phi^{(m)} = \arg \min_{\Phi \in \mathcal{C}_{\theta^{(m-1)}}} G(\Phi^{(m-1)}, \Phi). \quad (26)$$

We prove below in Lemma 3 that $g(r, s)$ satisfies two properties: a) $g(r, r) = h(r)$; and b) $g(s, s) \leq g(r, s)$ for all $r, s > 0$, and hence,

$$\begin{aligned} E_{\delta}(\Phi^{(m)}) &= G(\Phi^{(m)}, \Phi^{(m)}) \\ &\leq G(\Phi^{(m-1)}, \Phi^{(m)}) \\ &\leq G(\Phi^{(m-1)}, \Phi^{(m-1)}) \\ &= E_{\delta}(\Phi^{(m-1)}), \end{aligned}$$

where for the second inequality we use (26) and the fact that $\Phi^{(m-1)} \in \mathcal{C}_{\theta^{(m-1)}}$. The latter inclusion holds because we set $\theta^{(m-1)}$ based on the angles in $\Phi^{(m-1)}$.

Part (ii): Since

$$\|\tilde{\mathbf{p}}_i^{(m)} - \mathbf{q}_i\|^p \leq h(\|\tilde{\mathbf{p}}_i^{(m)} - \mathbf{q}_i\|) \leq E_{\delta}(\Phi^{(m)}) \leq E_{\delta}(\Phi^{(0)}),$$

we have that $\|\Phi^{(m)}\|_F^2 = \sum_i \|\tilde{\mathbf{p}}_i^{(m)}\|^2$ is bounded for all m and therefore has a convergent subsequence $\Phi^{(m_k)} \rightarrow \Phi^*$.

Part (iii): $B_j^{(m_k)}, C_j^{(m_k)}$ are defined as a linear functions (and hence continuous) in the entries of $A_j^{(m_k)}$ (see (8)) and therefore also converges to the similarity (resp., anti-similarity) parts of the limit map, i.e. $B_j^{(m_k)} \rightarrow B_j^*, C_j^{(m_k)} \rightarrow C_j^*$. We want to show that $A_j^* \in \mathcal{C}_{\theta_j^*}$, for some angle θ_j^* . Since $A_j^{(m_k)} \in \mathcal{C}_{\theta_j^{(m_k-1)}}$ it satisfies in particular (see (12))

$$\|C_j^{(m_k)}\|_\infty \leq \frac{K-1}{K+1} \frac{1}{2} \|B_j^{(m_k)}\|_F.$$

Taking the limit in both sides we get that $A_j^* \in \mathcal{C}_{\theta_j^*}$, where $\theta_j^* = \text{angle}(B_j^*)$. Denote $\theta^* = (\dots, \theta_j^*, \dots)$ and we get that $\Phi^* \in \mathcal{C}_{\theta^*}$. Note also that up to integer factor of 2π we have that $\theta_j^{(m_k)} \rightarrow \theta_j^*$, and therefore (with the correct 2π additions) $\theta^{(m_k)} \rightarrow \theta^*$, where we denote $\theta^{(m_k)} = (\dots, \theta_j^{(m_k)}, \dots)$.

Part (iv): Next we show that the limit mapping, Φ^* is a critical point of G over the limit set \mathcal{C}_{θ^*} . For this we will need to compare elements in the subsequence, $\Phi^{(m_k)}, \Phi^{(m_{k+1})}, \dots$ with their subsequent elements $\Phi^{(m_{k+1})}, \Phi^{(m_{k+1}+1)}, \dots$. For arbitrary $\Phi'_k \in \mathcal{C}_{\theta^{(m_k)}}$,

$$\begin{aligned} E_\delta(\Phi^{(m_{k+1})}) &\leq E_\delta(\Phi^{(m_k+1)}) \\ &= G(\Phi^{(m_k+1)}, \Phi^{(m_k+1)}) \\ &\leq G(\Phi^{(m_k)}, \Phi^{(m_k+1)}) \\ &\leq G(\Phi^{(m_k)}, \Phi'_k). \end{aligned}$$

Let $\Phi' \in \mathcal{C}_{\theta^*}$ be arbitrary. Since $\theta^{(m_k)} \rightarrow \theta^*$ we can choose $\Phi'_k \in \mathcal{C}_{\theta^{(m_k)}}$ such that $\Phi'_k \rightarrow \Phi'$. Now passing to the limit in the above inequality we get that for all $\Phi' \in \mathcal{C}_{\theta^*}$,

$$G(\Phi^*, \Phi') = E_\delta(\Phi^*) \leq G(\Phi^*, \Phi'). \quad (27)$$

The gradients of $E_\delta(\Phi)$, and $G(\Phi^*, \Phi)$ (as a function of the second variable again) w.r.t. the unknowns $\Phi = (\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2, \dots, \tilde{\mathbf{p}}_N)$ are:

$$\nabla_{\tilde{\mathbf{p}}_i} E_\delta(\Phi) = p(\tilde{\mathbf{p}}_i - \mathbf{q}_i) \left(\|\tilde{\mathbf{p}}_i - \mathbf{q}_i\|^2 + \delta \right)^{\frac{p}{2}-1} \quad (28)$$

$$\nabla_{\tilde{\mathbf{p}}_i} G(\Phi^*, \Phi) = p(\tilde{\mathbf{p}}_i - \mathbf{q}_i) \left(\|\tilde{\mathbf{p}}_i - \mathbf{q}_i\|^2 + \delta \right)^{\frac{p}{2}-1} \quad (29)$$

Hence $\nabla_{\tilde{\mathbf{p}}_i} E_\delta(\Phi^*) = \nabla_{\tilde{\mathbf{p}}_i} G(\Phi^*, \Phi^*)$.

Eq. (27) implies that Φ^* is a global minimum of $G(\Phi^*, \cdot)$ over the set \mathcal{C}_{θ^*} . Since this function ($G(\Phi^*, \Phi)$ as a function of Φ) is convex and the set of constraints $\Phi \in \mathcal{C}_{\theta^*}$ is convex, this implies in particular that

$$\langle \nabla_{\tilde{\mathbf{p}}_i} G(\Phi^*, \Phi^*), \Phi - \Phi^* \rangle \geq 0, \quad (30)$$

for all Φ satisfying $\Phi \in \mathcal{C}_{\theta^*}$. Therefore, we have

$$\langle \nabla_{\tilde{\mathbf{p}}_i} E_\delta(\Phi^*), \Phi - \Phi^* \rangle \geq 0,$$

for all Φ satisfying $\Phi \in \mathcal{C}_{\theta^*}$, and so Φ^* is a critical point of E_δ in \mathcal{C}_{θ^*} . \square

LEMMA 3. Let $h(r) = (r^2 + \delta)^{\frac{p}{2}}$, $g(r, s) = h(r) + \frac{p}{2}(r^2 + \delta)^{\frac{p}{2}-1}(s^2 - r^2)$. Then,

- (1) $g(r, s)$ is a convex quadratic function as a function of s .
- (2) $h(r) = g(r, r)$.
- (3) $g(s, s) \leq g(r, s)$ for all $r, s > 0$.

PROOF. 1 and 2 are clear. We will prove 3. It is enough to show

$$(s^2 + \delta)^{\frac{p}{2}} \leq (r^2 + \delta)^{\frac{p}{2}} + \frac{p}{2}(r^2 + \delta)^{\frac{p}{2}-1}(s^2 - r^2).$$

Let us multiply both sides by the positive number $(r^2 + \delta)^{1-\frac{p}{2}}$ and rearrange,

$$(r^2 + \delta)^{1-\frac{p}{2}}(s^2 + \delta)^{\frac{p}{2}} \leq (1 - \frac{p}{2})(r^2 + \delta) + \frac{p}{2}(s^2 + \delta),$$

and this follows from the standard inequality of weighted arithmetic versus geometric means. \square

ACKNOWLEDGMENTS

This research was supported in part by the European Research Council (ERC Starting Grant SurfComp, Grant No. 307754), U.S.-Israel Binational Science Foundation, Grant No. 2010331, by the Israel Science Foundation Grants No. 1284/12 and 764/10, I-CORE program of the Israel PBC and ISF (Grant No. 4/11), the Israeli Ministry of Science, and by the Citigroup Foundation.

REFERENCES

- BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLDMAN, D. B. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. on Graph.* **28**, 3.
- BELONGIE, S., MALIK, J., AND PUZICHA, J. 2002. Shape matching and object recognition using shape contexts. *TPAMI* **24**, 4, 509–522.
- BERG, A. C., BERG, T. L., AND MALIK, J. 2005. Shape matching and object recognition using low distortion correspondence. In *CVPR*. 26–33.
- BISSANTZ, N., DUMBGEN, L., MUNK, A., AND STRATMANN, B. 2009. Convergence analysis of generalized iteratively reweighted least squares algorithms on convex function spaces. *SIAM J. on Optimization* **19**, 4, 1828–1845.
- BOMMES, D., CAMPEN, M., EBKE, H.-C., ALLIEZ, P., AND KOBELT, L. 2013. Integer-grid maps for reliable quad meshing. *ACM Transactions on Graphics* **32**, 4.
- BOUAZIZ, S., TAGLIASACCHI, A., AND PAULY, M. 2013. Sparse iterative closest point. *Computer Graphics Forum (Symposium on Geometry Processing)* **32**, 5, 1–11.
- BRONSTEIN, A. M., BRONSTEIN, M. M., AND KIMMEL, R. 2006. Efficient computation of isometry-invariant distances between surfaces. *SIAM J. Scientific Computing* **28**, 5, 1812–1836.
- BROWN, M. AND LOWE, D. G. 2003. Recognising panoramas. In *ICCV*. Vol. 2.
- BROX, T. AND MALIK, J. 2011. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**, 3, 500–513.
- CHARTRAND, R. AND YIN, W. 2008. Iteratively reweighted algorithms for compressive sensing. In *ICASSP*. 3869–3872.
- CHUI, H. AND RANGARAJAN, A. 2003. A new point matching algorithm for non-rigid registration. *CVIU* **89**, 2, 114–141.
- DAUBECHIES, I., DEVORE, R., FORNASIER, M., AND GNTRK, C. S. I. 2010. Iteratively reweighted least squares minimization for sparse recovery. *Com. on Pure and Applied Mathematics* **LXIII**, 1–38.
- DUCHENNE, O., BACH, F., KWEON, I.-S., AND PONCE, J. 2011. A tensor-based algorithm for high-order graph matching. *TPAMI* **33**, 12, 2383–2395.
- ECKHARDT, U. 1980. Webers problem and weiszfelds algorithm in general spaces. *Mathematical Programming* **18**, 1, 186–196.

- FISCHLER, M. AND BOLLES, R. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Com. of the ACM **24**, 6, 381–395.
- FUNKHOUSER, T. AND SHILANE, P. 2006. Partial matching of 3d shapes with priority-driven search. In Proc. of the fourth Eurographics Symp. on Geometry Processing. 131–142.
- GELFAND, N., MITRA, N. J., GUIBAS, L. J., AND POTTMANN, H. 2005. Robust global registration. In Proc. of the third Eurographics Symp. on Geometry Processing.
- HACOHEN, Y., SHECHTMAN, E., GOLDMAN, D. B., AND LISCHINSKI, D. 2011. Non-rigid dense correspondence with applications for image enhancement. ACM Trans. Graph. **70**, 1–10.
- HEIDER, P., PIERRE-PIERRE, A., LI, R., AND GRIMM, C. 2011. Local shape descriptors, a survey and evaluation. In Eurographics Workshop on 3D Object Retrieval.
- HINTON, G., WILLIAMS, C., AND REVOW, M. 1993. Adaptive elastic models for hand-printed character recognition. NIPS, 512–512.
- JACOBSON, A., BARAN, I., POPOVIC, J., AND SORKINE, O. 2011. Bounded biharmonic weights for real-time deformation. ACM Trans. on Graph. **30/4**, 78, 1–78.
- JIAN, B., VEMURI, B., AND MARROQUIN, J. 2005. Robust nonrigid multimodal image registration using local frequency maps. In Information Processing in Medical Imaging. 289–301.
- LAWSON, C. L. 1961. Contributions to the theory of linear least maximum approximations. In PhD thesis, UCLA.
- LAZEBNIK, S., SCHMID, C., AND PONCE, J. 2004. Semi-local affine parts for object recognition. In BMVC. Vol. 2. 959–968.
- LAZEBNIK, S., SCHMID, C., AND PONCE, J. 2005. A maximum entropy framework for part-based texture and object recognition. In ICCV. Vol. 1. 832–838.
- LEORDEANU, M. AND HEBERT, M. 2005. A spectral technique for correspondence problems using pairwise constraints. In ICCV. Vol. 2. 1482–1489.
- LIPMAN, Y. 2012. Bounded distortion mapping spaces for triangular meshes. ACM Trans. Graph. **31**, 4, 108:1–108:13.
- LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. IJCV **60**, 2, 91–110.
- MONTAGNAT, J., DELINGETTE, H., AND AYACHE, N. 2001. A review of deformable surfaces: topology, geometry and deformation. Image and Vision Computing **19**, 14, 1023–1040.
- NEALEN, A., MULLER, M., KEISER, R., BOXERMAN, E., AND CARLSON, M. 2006. Physically based deformable models in computer graphics. Computer Graphics Forum **25**, 4, 809–836.
- SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: Exploring photo collections in 3d. In SIGGRAPH Conference Proceedings. ACM Press, New York, NY, USA, 835–846.
- STRECHA, C., VON HANSEN, W., GOOL, L. V., FUA, P., AND THOENNESSEN, U. 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In CVPR.
- TEVS, A., BOKELOH, M., WAND, M., SCHILLING, A., AND SEIDEL, H.-P. 2009. Isometric registration of ambiguous and partial data. In CVPR.
- TUYTELAARS, T. AND MIKOLAJCZYK, K. 2008. Local invariant feature detectors: a survey. Foundations and Trends in Computer Graphics and Vision **3**, 3, 177–280.
- VAN KAICK, O., ZHANG, H., HAMARNEH, G., AND COHEN-OR, D. 2011. A survey on shape correspondence. Computer Graphics Forum **30**, 6.
- VEDALDI, A. AND FULKERSON, B. Vlfat vision software.
- ZHENG, Y. AND DOERMANN, D. 2006. Robust point matching for non-rigid shapes by preserving local neighborhood structures. TPAMI **28**, 4, 643–649.

Received XXX; accepted XXX

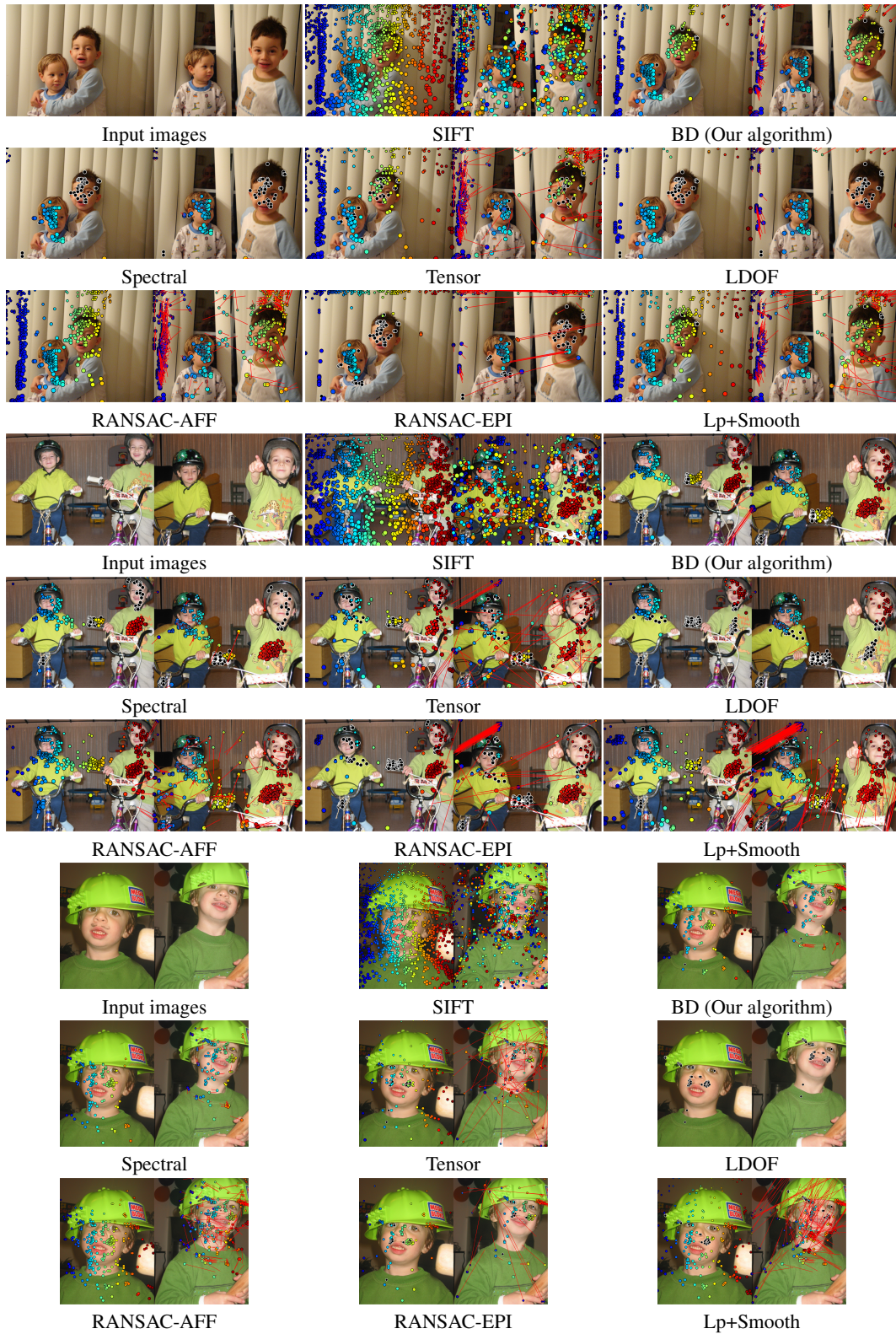


Fig. 5. Experiments with real images. Each set shows an image pair, SIFT matches used as input, and different methods for filtering the SIFT matches (BD is the method described this paper). Correspondences are shown with disks of the same size and color, with the color and size respectively varying linearly according to the horizontal and vertical position of points in the left image. On the right image in each pair, red lines show the deviation from ground truth locations marked by a user (red lines are not shown for the SIFT image to avoid clutter). Black disks depict good matches that were missed. ACM Transactions on Graphics, Vol. VV, No. N, Article XXX, Publication date: Month YYYY.



Fig. 6. Additional experiments with real images, see caption for Figure 5. Images taken from [Snavely et al. 2006; Strecha et al. 2008], used with the permission of the original authors.



Fig. 7. Additional experiments with real images, see caption for Figure 5. Images taken from [Lazebnik et al. 2004; 2005], used with the permission of the original authors.