Pattern Recognition Letters 32 (2011) 731–739

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Versatile spectral methods for point set matching

Alberto Silletti [a,*], Alessandro Abate [b,c], Jeffrey D. Axelrod [c], Claire J. Tomlin [d]

[a] Department of Information Engineering, University of Padova, Padova, Italy
[b] Delft Center for Systems and Control, TU Delft – Delft University of Technology, Delft, The Netherlands
[c] Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA
[d] Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA, USA

ARTICLE INFO

Article history:
Received 3 October 2009
Available online 5 December 2010
Communicated by R. Davies

Keywords:
Point set matching
Transformations and correspondences
Registration
Similarity metrics
Spectral methods
Kernel methods

ABSTRACT

This work is concerned with the problem of point set matching over features extracted from images. A novel approach to the problem is proposed which leverages different techniques from the literature. It combines a number of similarity metrics that quantify measures of correspondence between the two sets of features and introduces a non-iterative algorithm for feature matching based on spectral methods. The flexibility of the technique allows its straightforward application in a number of diverse scenarios, thus overcoming domain-specific limitations of known techniques. The proposed approach is tested in a number of heterogeneous case studies: of synthetic nature; drawn from experimental biological data; and taken from known benchmarks in computer vision.

1. Introduction

The general problem of point set matching is a fundamental topic in computer vision and is key for the task of registration of multiple images. The problem is defined over pairs of feature sets extracted from images and can be decomposed, as suggested in (Chui and Rangarajan, 2003), into two related sub-problems: that of *transformation* between the images, and that of *correspondence* between the features of the two images.

The first problem (transformation) is concerned with finding the mathematical relationship underlying the overall morphing between two successive frames, that is a mapping describing the transformation of the first image into the second frame (Ullman, 1979). This objective is relatively easily obtained when the underlying transformation is rigid: our approach instead does not assume that the transformation between the images is rigid.

The second issue (correspondence), which is the focus of the present contribution, deals with the task of finding quantitative matching between features associated with the two images, which are not necessarily related by a rigid transformation. In the literature this problem is studied either by working directly on the intensity maps associated with the two images (Ullman, 1979; Beauchemin and Barron, 1995), or by extracting correspondences between sets of features obtained from the images (Scott and Longuet-Higgins, 1991; Shapiro and Brady, 1992; Chui and Rangarajan, 2003). It is often the case that the two sets of features are composed of heterogeneously distributed points extracted from the original frames.

The contribution in (Huttenlocher and Rucklidge, 1993) employed a brute-force approach for the correspondence problem by exploiting the notion of Hausdorff distance. However the computational burden of the approach arising from combinatorial explosion can be significant. A popular, non-iterative technique employs spectral methods over the abstract structures composed of the feature points. Seminal contributions in this area are those in (Scott and Longuet-Higgins, 1991; Shapiro and Brady, 1992), which essentially differ on one major aspect. The former sets up a similarity metric based on *inter*-relationships between pairs of points taken from the two sets, and studies the spectral properties of a matrix that incorporates such a metric. The latter instead compares the spectral properties of two matrices, each of which is intended to describe the shape of the single image and its local relations (*intra-metrics* within the image). The technique is related to other approaches based on matrix spectral analysis for abstract weighted graph matching problems. Similar to the cited work in (Scott and Longuet-Higgins, 1991; Shapiro and Brady, 1992) and to (Umeyama, 1988), the presented technique is analytic. Alternatively, abstract graph matching can be studied with iterative schemes, as in (Blondel et al., 2004), which offset the computation time related to the scheme with a linear dependence on the graph size.

* Corresponding author. Tel.: +39 3286883343; fax: +39 0498277678.
E-mail addresses: alberto.silletti@unipd.it, silletti@dei.unipd.it (A. Silletti), a.abate@tudelft.nl (A. Abate), jaxelrod@stanford.edu (J.D. Axelrod), tomlin@eecs.berkeley.edu (C.J. Tomlin).

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.11.022

Both Scott and Longuet-Higgins (1991), Shapiro and Brady (1992) have known pitfalls and shortcomings. Scott and Longuet-Higgins (1991) presents limitations when the rotation or the scaling between the two images is large. To mitigate this deficiency, (Pilu, 1997) incorporates a neighborhood correlation measure in the *inter*-metrics. A spectral matching algorithm working along the lines of Scott and Longuet-Higgins (1991), which exploits intensity information from the raw underlying image, is proposed in (Srinivasan and Kankanhalli, 2003). As (Shapiro and Brady, 1992) argues, (Scott and Longuet-Higgins, 1991) suffers also from numerical instability. On the other hand, the approach in (Shapiro and Brady, 1992) performs poorly whenever the features corresponding to the main eigenmodes of either image undergo structural changes. Furthermore, the technique is not robust to point-jitter. Shapiro and Brady (1992) can be improved by employing kernel methods in the definition of the intra-metrics matrices (Shawe-Taylor and Cristianini, 2004). This is for instance implemented with a kernel PCA in (Wang and Hancock, 2004; Wang and Hancock, 2006), where the data are embedded in a higher-dimensional space through a Gaussian kernel. Along with not accommodating the presence of underlying images, both the approaches in (Scott and Longuet-Higgins, 1991) and in (Shapiro and Brady, 1992) do not incorporate information about feature connectivity or feature distinctiveness, which the original images may be endowed – (Sclaroff and Pentland, 1995) proposes an improvement to the correspondence procedure by incorporating outside information.

Other techniques employed in the literature study variants of the correspondence detection problem. Statistical methods (Wells, 1997) embed probability distributions over the point sets. The registration task can then be reframed as a maximum likelihood estimation procedure with coherence constraints (Myronenko et al., 2007), or as a distribution alignment problem (Jain and Vemuri, 2005). On the other hand, deterministic optimization-based approaches based on minimum least-squares problems (Chui and Rangarajan, 2003), or constrained global energy minimization techniques using objective functions quantifying geometric similarity and features coherence (Torresani et al., 2008) have been put forward to tackle the correspondence and matching problem. These techniques are intrinsically different in nature than the method discussed in this work in that (1) they possibly embed probabilistic information over the data sets, (2) they do not fully exploit the texture information underlying the images, and (3) they are iterative (as opposed to being one-shot), since they leverage optimization procedures (e.g., energy minimization or likelihood maximization). As such, while these techniques generally aim at solving the same problem, they are methodologically and structurally different than the approach discussed in the present contribution.

The purpose of this work is to propose a novel approach for the general problem of point set matching. With reference to the earlier literature review, the approach (1) leverages and combines the *intra*- and *inter*-information obtained from the pair of feature sets; (2) exploits the use of spectral techniques over the feature sets – the proposed method is non iterative and achieves a computational efficiency; (3) embeds additional information deriving from the possible knowledge of an existing graphical structure over the single features set, as well as from the presence of underlying raw images. Such information is included by exploiting structural properties of kernel matrices.

As in (Scott and Longuet-Higgins, 1991), we make use of a *pairing matrix*, which relates pairs of points taken from the two sets. In addition, we allow the elements of the pairing matrix to depend on a combination of possible metrics, each of which is defined over the two sets of points. This "library" of metrics represents a set of different possible measures that quantify attributes of similarity between the pair of feature sets under study, and can be extended based on the problem under study. For example, one metric introduces a notion of adjacency between the spatial coordinates of pairs of points, as proposed in (Scott and Longuet-Higgins, 1991); two related metrics compare the eigenmodes between intra-matrices defined for each of the two feature sets, as in (Shapiro and Brady, 1992); an additional metric exploits the presence of a graphical structure underlying the points in each image; a final metric instead uses the information coming from actual raw images underlying the feature points. Thanks to structural properties of kernel matrices, such measures are combined into a single pairing matrix as the multiplication of the single pairing matrices. This new pairing matrix accommodates different and heterogeneous information: such versatility allows the application of the approach to diverse data sets.

The technique performs robustly on a variety of application domains, by automatically adapting the set of useful metrics to the particular case under study. Specifically, the multiplicative structure of the pairing matrix ensures that if a particular metric is not discriminative for a pair of images it is automatically overridden by the other more relevant metrics. The procedure is tested over a few case studies drawn from different domains, in order to fairly assess its performance with regard to the following major characteristic benchmarks: (1) the presence of outliers (i.e., the emergence or the deletion of a subset of the features) in each image due to structural changes; (2) substantial rotations and translations between the pair of images; and (3) random position jitter.

The contribution unfolds as follows. Section 2 introduces the theoretical concepts and delineates the main technical aspects underlying the proposed technique. In particular, Sections 2.1 and 2.2 discuss the notion of pairing matrix and its use for the point set matching problem, Section 2.3 introduces a library of different possible metrics to be adaptably embedded within the pairing matrix, and Section 2.4 derives the algorithmic complexity of the proposed approach. Section 3 covers a number of case studies, where the technique proposed in this work is tested and benchmarked. The examples are both synthetic (Section 3.2) and drawn from biological experimental data (Section 3.3), as well as taken from known benchmarks in the literature (http://vasc.ri.cmu.edu/idb/html/motion/house/index.html, Section 3.4). Section 4 concludes the work and delineates future research directions.

2. Theory

This Section discusses the theory that underpins the proposed point set matching procedure.

2.1. The pairing matrix for point set matching

Consider two sets of points $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$, with elements lying in $\mathbb{R}^d$ for some finite d. As in (Scott and Longuet-Higgins, 1991), we introduce a *pairing matrix* $\mathcal{Z} \in \mathbb{R}^{m \times n}$, with entries $z_{ij} \in \mathbb{R}$. Each element z_{ij} is intended to express a measure of similarity between point $x_i \in X$ and $y_j \in Y$. The definition of the elements in $\mathcal{Z}$ is elaborated in Sections 2.2 and 2.3. Given a matrix $\mathcal{Z}$, the selection of pairs of matching points from the two sets is performed after a normalization procedure, and is described in the following.

The matrix $\mathcal{Z}$ is preprocessed by computing its singular-value decomposition

$$\mathcal{Z} = TDU, \tag{1}$$

where $D \in \mathbb{R}^{m \times n}$, whereas T, U are properly-sized orthogonal matrices. We replace diagonal elements d_{pp} of D, $p = 1, 2, \ldots, \min\{m, n\}$, with identity constants, which yields

$$\widetilde{\mathcal{Z}} = TEU, \tag{2}$$

where $e_{ij} = d_{ij}$, $i \neq j$, $i = 1,\ldots,m$, $j = 1,\ldots,n$, $e_{pp} = 1$, $p = 1,\ldots,\min\{m,n\}$. This technique is generally known as *whitening*. The matrix $\widetilde{\mathcal{Z}}$ is orthogonal and is the one which maximizes the trace of $\widetilde{\mathcal{Z}}^T\mathcal{Z}$ (Scott and Longuet-Higgins, 1991). Furthermore, the largest elements in $\widetilde{\mathcal{Z}}$ correspond to candidate matching pairs as follows: the pair (x_i,y_j) is matched if and only if z_{ij} is the largest element both of row i and of column j. This strong correspondence implies a "mutual consent" to the match: indeed, if z_{ij} is the largest element of row i but not of column j, point x_i is similar to y_j, but not the contrary. As such, the pair (x_i,y_j) is not a valid match.

If one considers each row i as an n dimensional vector r_i, then $\mathcal{Z}$ is a map from point x_i into vector r_i. Ideally, a pairing matrix $\mathcal{Z}$ should be sparse, with a single non-zero element per row and linearly independent rows. In such a case, r_i coincides with one of the axis and as such is the farthest possible from any other vector row. More generally, if a row vector r_i is close to the axis $e_j \in \mathbb{R}^n$, then it is likely for the pair (x_i,y_j) to be a match. However, if two vectors r_{i_1} and r_{i_2}, $i_1,\ i_2 \in \{1,\ldots,m\}$, are adjacent to the same axis e_j, then the corresponding points x_{i_1} and x_{i_2} "compete" for the match with y_j. Setting the singular values of $\mathcal{Z}$ to be unitary corresponds to a spatial outspread of its row vectors, thus alleviating such potential conflicts. Fig. 1 displays an instance of $\mathcal{Z}$ and $\widetilde{\mathcal{Z}}$ matrices corresponding to sets of cardinality three. In this example where $m = n$, the above operation corresponds to the normalization the volume of the associated prism.

Unlike other iterative approaches (Chui and Rangarajan, 2003; Torresani et al., 2008), the method yields the matching with a single-step calculation. The formal algorithmic complexity of the approach is derived in Section 2.4.

2.2. Flexible design of the pairing matrix

As previously mentioned, the elements z_{ij} of $\mathcal{Z}$ represent a measure of similarity between point x_i and point y_j. A kernel function $\mathcal{K} : X \times Y \to \mathbb{R}$ is used to define these elements as an inner product in a (possibly infinite) dimensional space (Shawe-Taylor and Cristianini, 2004), so that

$$z_{ij} = \mathcal{K}(x_i,y_j) = \langle \phi(x_i), \phi(y_j) \rangle \tag{3}$$

and

$$\mathcal{Z} = \begin{bmatrix} z_{11} & \ldots & z_{1n} \\ \vdots & \ddots & \vdots \\ z_{m1} & \ldots & z_{mn} \end{bmatrix} = \begin{bmatrix} \mathcal{K}(x_1,y_1) & \ldots & \mathcal{K}(x_1,y_n) \\ \vdots & \ddots & \vdots \\ \mathcal{K}(x_m,y_1) & \ldots & \mathcal{K}(x_m,y_n) \end{bmatrix}. \tag{4}$$

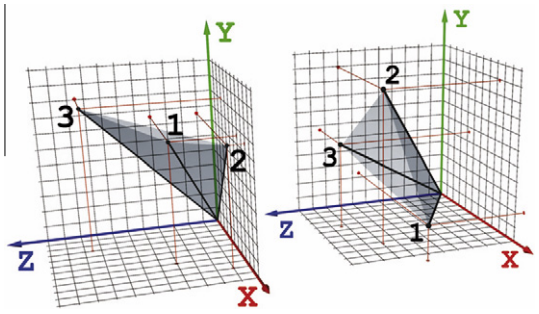


Fig. 1. Graphical comparison of pairing matrix $\mathcal{Z}$ (left) and its normalized version $\widetilde{\mathcal{Z}}$ (right), where $\mathcal{Z} = \begin{bmatrix} 0.7000 & 0.7500 & 0.4000 \\ 0.9000 & 0.9500 & 0.1000 \\ 0.3000 & 0.9000 & 0.8500 \end{bmatrix}$ and $\widetilde{\mathcal{Z}} = \begin{bmatrix} 0.8519 & 0.1527 & 0.5010 \\ 0.4249 & 0.7607 & 0.4907 \\ 0.3062 & 0.6309 & 0.7129 \end{bmatrix}$

Notice that the row vectors r_1, r_2, r_3 of $\widetilde{\mathcal{Z}}$ do not cluster and better spread in space. The pairing matrix $\mathcal{Z}$ yields one pair (2,2), whereas $\widetilde{\mathcal{Z}}$ yields the pairs (1,1), (2,2), (3,3).

As a special case of (3), one can select $\mathcal{K}(x_i,y_j) = \langle x_i,y_j \rangle$, whereby $\phi(\cdot) = \mathrm{id}(\cdot)$ is the identity function. In general, given a proper kernel function $\mathcal{K}$, Mercer's Theorem (Shawe-Taylor and Cristianini, 2004, Thm. 3.13) guarantees the existence of the embedding function $\phi(\cdot)$, though it may not be explicitly known.

Given a (possibly infinite) set of kernel functions $\mathcal{K}_1,\ldots,\mathcal{K}_t$, $t \in \mathbb{N}$, the closure property (Shawe-Taylor and Cristianini, 2004, Prop.3.22) warrants that

$$\mathcal{K}^s(x_i,y_j) = \sum_{k=1}^{t} \mathcal{K}_k(x_i,y_j) \quad \text{and} \quad \mathcal{K}^p(x_i,y_j) = \prod_{k=1}^{t} \mathcal{K}_k(x_i,y_j) \tag{5}$$

are also kernel functions. Hence, considering pairing matrices $\mathcal{Z}_1,\ldots,\mathcal{Z}_t$ built on the kernel functions $\mathcal{K}_1,\ldots,\mathcal{K}_t$ as in (4), the new matrices

$$\mathcal{Z}^s = \sum_{k=1}^{t} \mathcal{Z}_k \quad \text{and} \quad \mathcal{Z}^p = \prod_{k=1}^{t} \mathcal{Z}_k \tag{6}$$

are still valid kernel-based pairing matrices. (Here the symbols $\sum$ and $\prod$ represent point-wise sum and multiplication.)

For any legitimate choice of $\mathcal{K}$, the matrix $\mathcal{Z}$ relates to the notion of Gram matrix (Shawe-Taylor and Cristianini, 2004), which connects with a wealth of literature on kernel methods for pattern analysis. However, since in the present instance $\mathcal{Z}$ is applied over pairs of points extracted from two different sets (rather than from the very same set), it does not formally belong to this class of matrices.

The kernels $\mathcal{K}_k$ are built around metrics, or distances d_k, which are suitable for the matching problem. We select a Gaussian as the kernel function:

$$\mathcal{K}_k(x_i,y_j) = e^{-\frac{d_k^2(x_i,y_j)}{\sigma_k^2}}, \tag{7}$$

where σ_k is a parameter of choice that controls the degree of interaction between the two feature points.

Next, we introduce a library of possible metrics, which we will apply in a number of heterogeneous case studies.

2.3. A library of metrics

Let us consider the following set of metrics $d_k : X \times Y \to \mathbb{R}$, $k = 1,\ldots,5$:

Metric	Definition
$d_1(x_i,y_j)$	$\|x_i - y_j\|_p$, $p > 0$
$d_2(x_i,y_j)$	$\cos(\mathtt{m}(i), \mathtt{m}(j))$
$d_3(x_i,y_j)$	$\cos(\tilde{\mathtt{m}}(i), \tilde{\mathtt{m}}(j))$
$d_4(x_i,y_j)$	$\lvert\mathtt{d}(x_i) - \mathtt{d}(y_j)\rvert$
$d_5(x_i,y_j)$	$\mathtt{td}(x_i,y_j)$

The metrics are further discussed in the following.

1. d_1 is a p-norm distance between pairs of points taken from the two sets. The present case studies consider the Euclidean norm ($p = 2$).
2. d_2 is a measure of the distance of the *ith mode* $\mathtt{m}(i)$ and the *j*th *mode* $\mathtt{m}(j)$ associated to each of the two sets. As suggested in (Shapiro and Brady, 1992), the mode $\mathtt{m}$ of a point set is computed as follows: first, a square proximity matrix defined according to the intra-distances between the features of the image is introduced; it is successively diagonalized and its first $\min\{m,n\}$ eigenvectors, sorted according to the largest eigenvalues, are regarded as its modes. The distance between the modes $\mathtt{m}(i)$ and $\mathtt{m}(j)$ of the two graphs is then computed as their cosine.

3. This metric is valid if the sets X, Y are embedded with a graphical structure. The distance d_3 is characterized analogously as d_2, however the modes $\tilde{\mathtt{m}}$ are computed by considering a proximity matrix that has non-zero elements only if the corresponding pair of vertices are connected by an existing edge in the graph.
4. The metric d_4 is defined as the absolute value of the difference between the graphical degree of a pair of points, taken respectively from X and from Y. The function $\mathtt{d}$ is the degree of a node in a graph. As for d_3, this metric is valid if the sets X, Y are embedded with a graphical structure. However, if no graphical structure is present over the sets X and Y, then one such graph can be induced artificially: for instance, edges can be created between pairs of points if their distance is less than a adjustable threshold.
5. Given an image underlying the point sets, the metric d_5, defined by a function $\mathtt{td}$, computes the *texture difference* between a neighborhood of $x_i \in X$ and one of $y_j \in Y$. Here the points x_i, y_j are intended as features of the corresponding images.
   More formally, let us consider the finite discrete domain $L \subset \mathbb{Z}^2$ made up of the pixels of the two-dimensional image frame. Given a point $z \in \mathbb{R}^2$, we define a neighborhood $\mathcal{N}(z,\delta) \subset \mathbb{Z}^2$ as the set of pixels of the image with centroid lying within a radius $\delta > 0$ from z. A function $\mathcal{I} : L \to \mathbb{R}^+$ specifies the *intensity* of the image over its domain.
   Let us now consider the two images L_x, L_y underlying X, Y. Given a radius δ of choice (which value may be related to σ_5 in (7)), the function $\mathtt{td}(z_1,z_2)$ computes the absolute value of the difference between the intensities of the pixels in the neighborhoods of points z_1 and z_2:

$$\mathtt{td}(z_1,z_2) = \sum_{p_1 \in \mathcal{N}(z_1,\delta)\cap L_x} \; \sum_{p_2 \in \mathcal{N}(z_2,\delta)\cap L_y} |\mathcal{I}(p_1) - \mathcal{I}(p_2)|.$$

   This approach is related to a similar procedure used in (Srinivasan and Kankanhalli, 2003).

The metrics $d_1,\ldots,d_5$ used in kernel functions (7) are then combined through a product into a pairing matrix $\mathcal{Z}$ as in (5) for the matching procedure. The multiplicative structure of the pairing matrix ensures that if a particular metric is not discriminative for a pair of images it is automatically overridden by other more relevant metrics. This versatile feature allows the automatic adaptation of the set of useful metrics to the particular case under study.

2.4. Algorithmic complexity

Let us start assuming that we are given t kernel functions $\mathcal{K}_k,\ k = 1,\ldots,t$, as in (7), which are used in the matching procedure. The procedure is made up of four sequential steps that critically determine its algorithmic complexity.

Firstly, we build $\mathcal{Z}$ as a point-wise multiplication of the t kernel functions $\mathcal{K}_k,\ k = 1,\ldots,t$, according to (6). This involves $d \cdot t \cdot m \cdot n$ operations, where m and n are the number of features respectively in X and Y, and d is the dimensionality of their spatial component. We expect that d, t are constants that are smaller than m, n, which results in a $\mathcal{O}(nm)$ computational complexity.

Secondly, we compute $\widetilde{\mathcal{Z}}$ using a singular value decomposition procedure as in (Golub and Reinsch, 1970): accordingly, the computation of U, V and D requires $4m^2n + 8mn^2 + 9n^3$ operations.

Thereafter, the whitening procedure is completed by replacing the diagonal elements of D, which easily results in $\mathcal{O}(\min\{m,n\})$ computational complexity.

Finally, the matching procedure on $\widetilde{\mathcal{Z}}$ requires finding the minimum of each row and then scanning the corresponding column to confirm or discard the matching. This entails in $\binom{\min\{m,n\}}{2}$ operations, which is $\mathcal{O}((\min\{m,n\})^2)$.

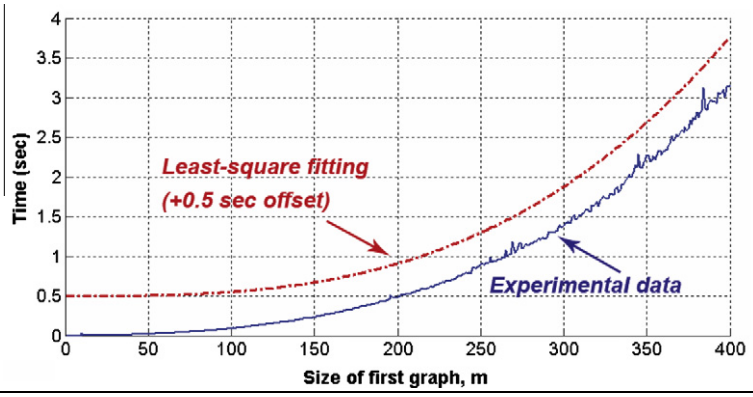
The overall computational complexity thus easily add up being $\mathcal{O}(m^2n + mn^2 + n^3)$. Intuitively, we can state that the computation cost grows with the third power of the size of $|X|$ or $|Y|$.

In practice, we have found the computational cost not to be a burden, even using whitening. All the case studied described in Section 3, which are by no means of trivial dimensionality or setup, have resulted in real-time executions. In other words, whilst the computed polynomial complexity holds, we have experienced that the hidden constants in $\mathcal{O}(m^2n + mn^2 + n^3)$ to be quite limited in value. For instance, Table 1 shows the computational complexity of the matching procedure for the synthetic graph matching test case in Section 3.2, run using $\mathbb{P}_e$, $\mathbb{P}_v$ and $\mathbb{M}_v = 15.00\%$: for this test, we have found that the complexity $\mathcal{O}(m^2n + mn^2 + n^3)$ can be further computed to be equal to $2.1 \times 10^{-7}\,\mathrm{m}^3$, where m is the size of the first graph (here $n \leqslant m$). On the side, notice that one can implement a much faster algorithm, albeit at a loss in reliability, by avoiding the whitening of $\mathcal{Z}$ (second and third step above). This obtains a computational cost of $\mathcal{O}(mn)$.

3. Case studies

This Section develops a number of case studies to test the methodology introduced in Section 2. We show that the particular

Table 1
Computational complexity for the synthetic graph test case of Section 3.2, tested on a Linux Ubuntu machine, AMD Mobile Sempron 3600+ processor with 2 GB memory. The test was built in Matlab *r2009b*©. To reduce noise, for each size m of the graph we run 10 trials and averaged the outcomes. The table shows the experimental result (blue) as well the least square analytical fitting (red), obtained with a multiplicative constant 2.1×10^{-7} and here displayed with a 0.5 s offset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



Graph size, m	Computational Time
50 nodes	0.022 sec
75 nodes	0.047 sec
100 nodes	0.089 sec
125 nodes	0.154 sec
150 nodes	0.231 sec
175 nodes	0.347 sec
200 nodes	0.472 sec
250 nodes	0.894 sec
300 nodes	1.396 sec
350 nodes	2.277 sec
400 nodes	3.157 sec

instance under study dictates what subset of metrics ought to be selected to form the pairing matrix utilized for the matching procedure, in order to optimize its outcomes. The software and the hardware for all the case studies are specified in Table 1.

3.1. Performance assessment of the point set matching procedure

Consider the two sets of feature points X and Y, with cardinality m and n respectively. The outcomes of the point set matching procedure is compared with the ground truth. The ground truth is known for the synthetic case studies, whereas for the case studies based on real images it is directly assessed over the data sets by independent and unbiased observers. Let us introduce the following four entities:

1. $X_{tm} \subseteq X$, $Y_{tm} \subseteq Y$ are the two sets of feature points that are correctly matched, of cardinality respectively m_{tm}, n_{tm};
2. $X_{ts} \subseteq X$, $Y_{ts} \subseteq Y$ are the two sets of feature points that are correctly left un-matched, of cardinality respectively m_{ts}, n_{ts};
3. $X_{fm} \subseteq X$, $Y_{fm} \subseteq Y$ are the two sets of feature points that are wrongly matched, of cardinality respectively m_{fm}, n_{fm};
4. $X_{fs} \subseteq X$, $Y_{fs} \subseteq Y$ are the two sets of feature points that are wrongly left un-matched, of cardinality respectively m_{fs}, n_{fs}.

Notice, as intuitive, that $m_{tm} + m_{ts} + m_{fm} + m_{fs} = m$ and that $n_{tm} + n_{ts} + n_{fm} + n_{fs} = n$. We define as percentages, over both sets of points, the following quantities:

1. *true matches,* as the ratio of feature points that are correctly matched, i.e. $\frac{m_{tm}+n_{tm}}{m+n}$;
2. *true singles,* as the ratio of feature points that are correctly left un-matched, i.e. $\frac{m_{ts}+n_{ts}}{m+n}$;
3. *false matches,* as the ratio of feature points that are wrongly matched, i.e. $\frac{m_{fm}+n_{fm}}{m+n}$;
4. *false singles,* as the ratio of feature points that are wrongly left un-matched, i.e. $\frac{m_{fs}+n_{fs}}{m+n}$.

The outcome of the case studies will be evaluated according to the introduced quality measures.

3.2. Synthetic graph matching

We consider a graphical structure $G = (V, E)$ with two dimensional spatial components, which are constrained to lie within the unit square in the positive quadrant $[0,1]^2$. The cardinality of the set of nodes V is equal to 50 and their spatial components are generated uniformly at random within the specified domain. The edge set E is created between pairs of nodes in V according to a Bernoulli distribution with mean equal to 0.5, however the edges that are longer than a specified threshold (0.25) are discarded. Fig. 2 shows an example of a graph.

The graph is then morphed into a new structure $\widetilde{G} = (\widetilde{V}, \widetilde{E})$, according to the following procedure:

1. The set $\widetilde{E} \subseteq E$ is generated from E by discarding each edge according to a Bernoulli probability distribution with mean $\mathbb{P}_e$;
2. The set $\widetilde{V} \subseteq V$ is generated from V by discarding each edge according to a Bernoulli probability distribution with mean $\mathbb{P}_v$, and additionally by eliminating the residual edges that connect to vertices in $E \setminus \widetilde{E}$;
3. The spatial components associated to the elements in $\widetilde{E}$ are obtained from those belonging to the corresponding elements in E by perturbing them with the addition of a random variable that is uniformly distributed within the square $\frac{\mathbb{M}_v}{2}[-1,1]$. In other words, the original coordinates are subjected to a uniform perturbation that amounts to $\mathbb{M}_v\%$ of their maximum possible value.

The matching procedure proposed in Section 2 is tested on a cohort of pairs of graphs $(G, \widetilde{G})$, parameterized by the input configuration $(\mathbb{P}_e, \mathbb{P}_v, \mathbb{M}_v)$ used to generate them: for each combination $(\mathbb{P}_e, \mathbb{P}_v, \mathbb{M}_v)$ of perturbation parameters we average the outcomes of 2000 simulations.

Figs. 2 and 3 represent respectively a single pair of test graphs $(\mathbb{P}_e = \mathbb{P}_v = \mathbb{M}_v = 15.00\%)$, and the outcomes of the matching procedure.

We have employed the first four metrics d_1, d_2, d_3 and d_4. The metric d_5 is not employed, since the artificial graphs have no underlying physical image that can be exploited for the matching procedure. Table 2 reports the results for each configuration of the perturbation parameters $(\mathbb{P}_e, \mathbb{P}_v, \mathbb{M}_v)$ are the average of the 2000 simulations. The experiments are divided in five batches: in the first, we uniformly modify the three perturbation parameters; in the following four, we fix two of the three parameters and modify the remaining one by using values that match those of the first batch of experiments. The monotonicity of the performance outputs of the algorithm with respect to the perturbation level provides evidence of the consistency of the procedure (see first set of simulations). The results of the last four groups of simulations "lie within" those of the first batch (the comparison ought to be

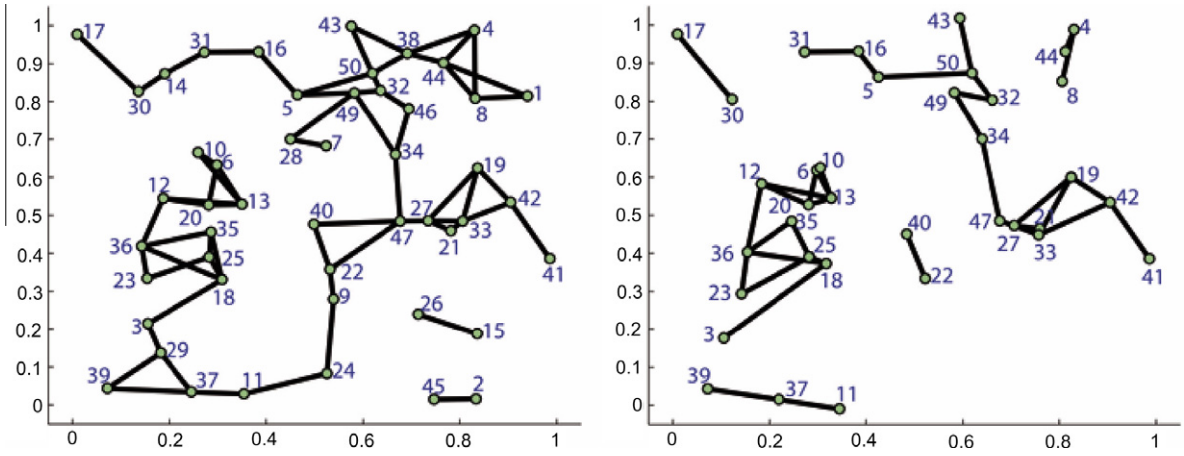


Fig. 2. Synthetic graph generation and perturbation. The original graph (left) contains 50 nodes. The perturbed graph (right) was generated using $\mathbb{P}_e = \mathbb{P}_v = \mathbb{M}_v = 15.00\%$. The blue labels mark the nodes and provide the ground truth correspondence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

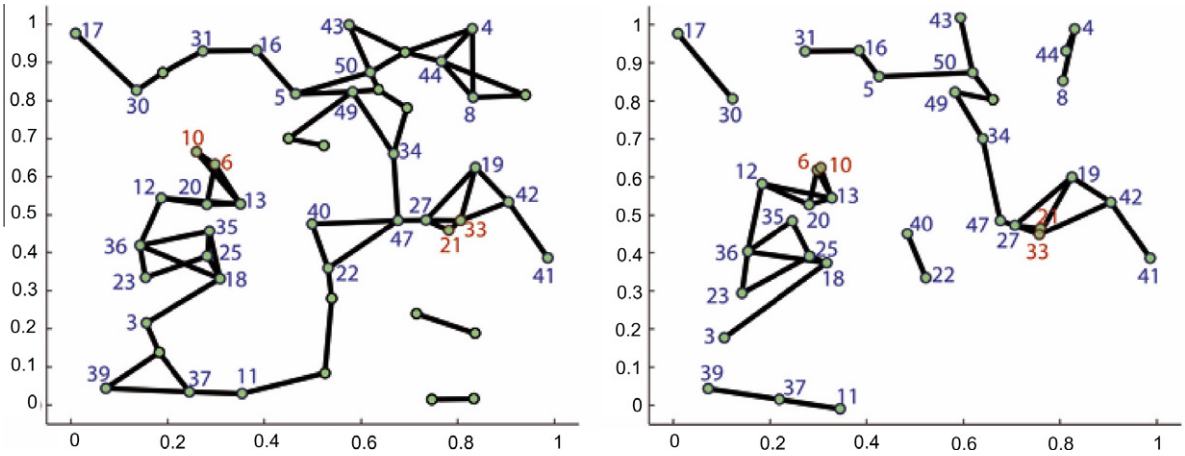


Fig. 3. Outcomes of the matching procedure for the synthetic graphs of Fig. 2. Here the blue labels indicate correctly matched nodes (true matches), whereas red labels shows the wrongly matched ones. Nodes without labels are correct single nodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Outcomes of the matching procedure tested on sets of randomly generated and successively perturbed graphs. We have run 2000 simulation for each configuration of perturbation parameters and reported the average of the outcomes. For each of the 2000 simulations we have first generated a graph, then perturbed it. The perturbation level is tuned via three parameters: $\mathbb{P}_e$, the probability that an edge is erased from the original graph; $\mathbb{P}_v$, the probability that a vertex is eliminated from the original graph; $\mathbb{M}_v$, the level of spatial perturbation applied to a vertex of the original graph.

Perturbation			Output Performance			
$\mathbb{P}_e$ (%)	$\mathbb{P}_v$ (%)	$\mathbb{M}_v$ (%)	True matches (%)	True singles (%)	False matches (%)	False singles (%)
15.00	15.00	15.00	71.01	7.50	19.37	3.12
12.00	12.00	12.00	79.83	6.66	12.07	1.44
9.00	9.00	9.00	87.65	4.37	7.10	0.88
6.00	6.00	6.00	93.06	3.38	3.36	0.20
3.00	3.00	3.00	97.09	1.82	1.04	0.05
12.00	15.00	15.00	72.40	7.21	18.59	2.80
15.00	12.00	15.00	73.17	6.05	18.64	2.14
15.00	15.00	12.00	78.07	8.94	12.19	1.90
9.00	12.00	12.00	80.29	6.71	11.41	1.59
12.00	9.00	12.00	82.31	4.32	12.05	1.32
12.00	12.00	9.00	84.49	6.64	7.80	1.07
6.00	9.00	9.00	87.86	5.62	6.03	0.49
9.00	6.00	9.00	89.31	3.32	6.72	0.65
9.00	9.00	6.00	90.86	5.23	3.42	0.59
3.00	6.00	6.00	93.26	3.60	2.78	0.36
6.00	3.00	6.00	95.17	1.68	2.99	0.16
6.00	6.00	3.00	94.86	3.70	1.20	0.24

done by looking at the accrued true and false value pairs). By comparing the third result of each group of experiment with the first two, one can observe that the elimination of edges or vertices affects the quality of the outcomes more than the perturbation of the spatial coordinates of the vertices. This is despite the fact that spatial perturbations can result in feature crossover. Furthermore, as intuitive, the elimination of an edge (first result of the last four groups) affects the results more than that of a vertex (second result).

3.3. Image registration of biological data: the Drosophila wing case study

This experimental study aims at matching two network structures extracted from biological data.[1] Each network describes the cellular epithelium of a wing of *Drosophila melanogaster*, the common fruit fly. The experimental data are obtained with confocal microscopy techniques a few hours after puparium formation. It is of interest for the developmental biologist to have access to quantitative data relating to the network structure of the epithelium. The graphical structure is extracted from single frames that belong to time-lapse movies of the epithelium. The details of the computer vision technique used to extract the network from a single frame are formally presented in (Silletti et al., 2009). Along with the collection of the graphical structures corresponding to each frame, it is important to match the networks extracted from pairs of frames that are successive in time. This procedure is also known as the *registration* of the frames of the movie.

The experimental data consist of 50 frames consecutive in time, corresponding to 49 pairs of images. Fig. 4 shows frames (frame 22 and 23) taken from the wet lab experimental data.

For the instance under study, we have employed the metrics d_1 and d_5. The use of d_5 is dictated by the availability of an actual image containing meaningful information for the matching. The *intra-metrics* d_2, d_3 and d_4 are discarded, which is explained by observing the similarity of the neighborhood structure for most of the nodes in the graph. In other words, if most of the internal nodes have a

[1] The images have been provided by the Axelrod Lab, at the Department of Pathology, Stanford University School of Medicine, Stanford, USA. Members of the Lab have also contributed in the interpretation of the outcomes of the registration procedure.

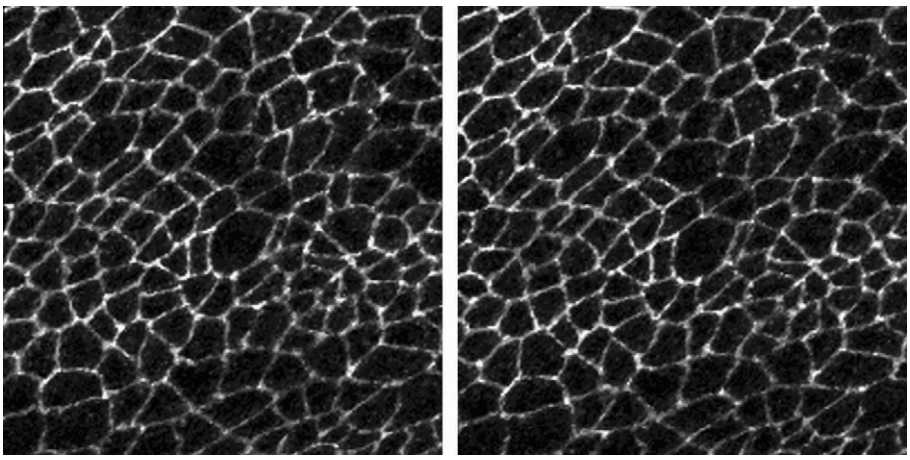


Fig. 4. Frames 22 and 23 considered for the matching procedure. The images are part of a 40 frame movie and refer to a section of the epithelium of the *Drosophila melanogaster* wing. The polygonal structures are 2-d sections of the epithelial cells.

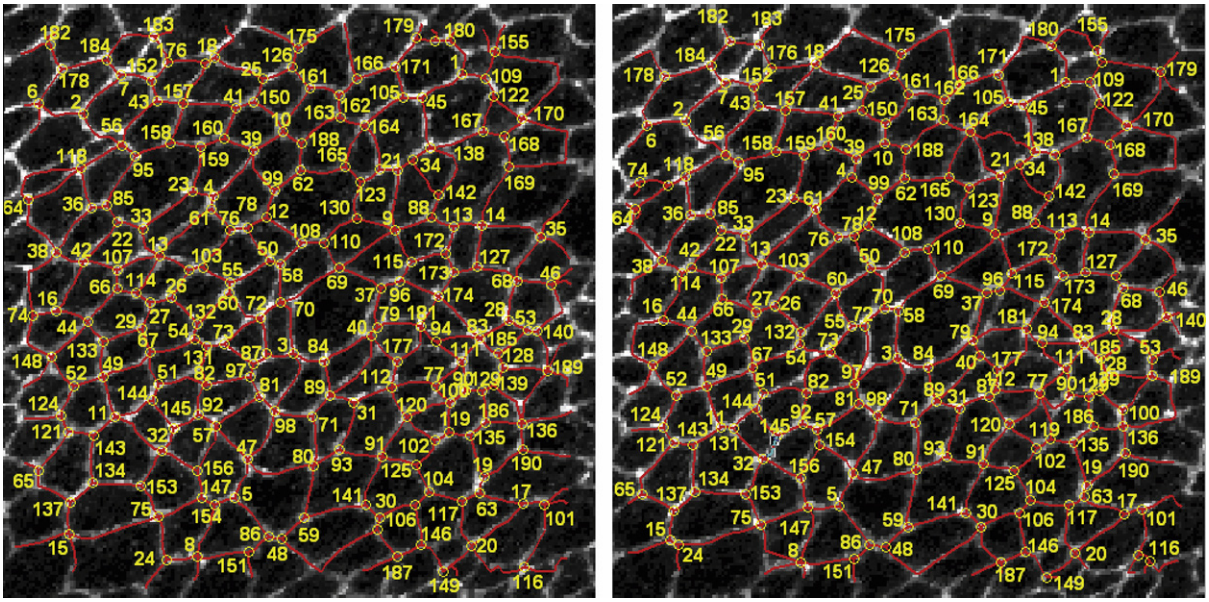


Fig. 5. The matching procedure applied to the networks of Fig. 4. The yellow labels over the nodes correspond to matches (both correct and wrong ones). The unlabelled nodes are single nodes (both correct and wrong ones). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

similar number of connected edges, then the information provided by the metric d_4 is redundant. Similar considerations hold for the metrics d_2, d_3.

Fig. 5 displays the graphical structures extracted from the frames in Fig. 4, and labeled with the outcome of the matching procedure. Unlike the study in Section 3.2, the ground truth has been provided by manual observation from an independent and unbiased observer. Table 3 displays the outcomes of the procedure, averaged over the 49 tests. The results appear to be rather good when compared to others in the biology literature (Ma et al., 2008), especially given the complexity of the structures and of the dynamics under study (cells both appear do to divide and to exit the epithelium, the frames are subject to translation, and the images are quite noisy – the last two are known issues for algorithms as in (Scott and Longuet-Higgins, 1991)). (See Fig. 6)

Table 3
Outcomes of the matching procedure tested on 49 pairs of networks extracted from 50 successive frames of a movie. The results are averages over the 49 tests. The movie refers to the morphogenesis and the dynamics of a section of the wing of *Drosophila melanogaster*.

Output performance			
True matches	True singles	False matches	False singles
88.23%	3.61%	7.92%	0.24%

3.4. Point set matching over a literature benchmark: the "CMU House"

We have finally tested our procedure on a known benchmark from the computer vision literature, known as the "CMU House" (http://vasc.ri.cmu.edu/idb/html/motion/house/index.html). This benchmark contains a set of 110 pictures of a toy house, taken over a black background. We have extracted a set of features from each image by applying a corner detector (Xiao and Nelson, 2008). The obtained sets have a cardinality that is very similar to the sets used in (Carcassoni and Hancock, 2003; Wang and Hancock, 2006) for the same benchmark, which allows for a fair comparison with those results.

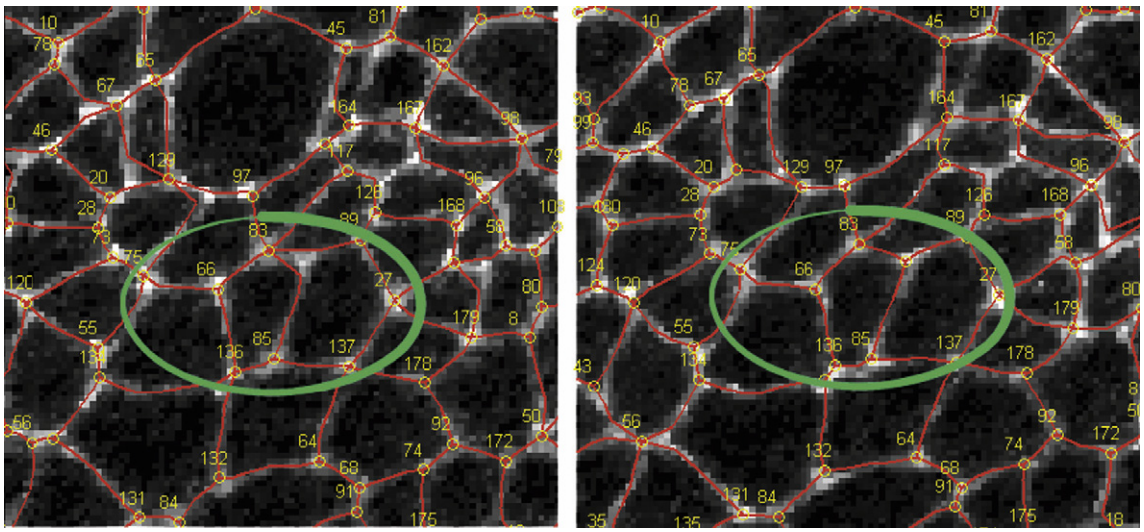


Fig. 6. A particular of the matching procedure from Fig. 5, involving a topological change. The green circle highlights a difficult match that is correctly resolved. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We have tested our algorithm on two experimental setups. Firstly, we have matched the 109 sequential pairs of images (1–2,2–3,...,109–110). Fig. 7 displays the output of one such pairing: the green labels are obtained from the matching procedure. Secondly, we have matched 109 pairs of distinct images, randomly chosen from the set. Fig. 8 displays the output of one such matching: in green are the correct labels obtained from the matching procedure, whereas in red are the wrong outcomes.

The first study is meant to test the robustness of the method with respect to positional jitter, while the second targets the per-

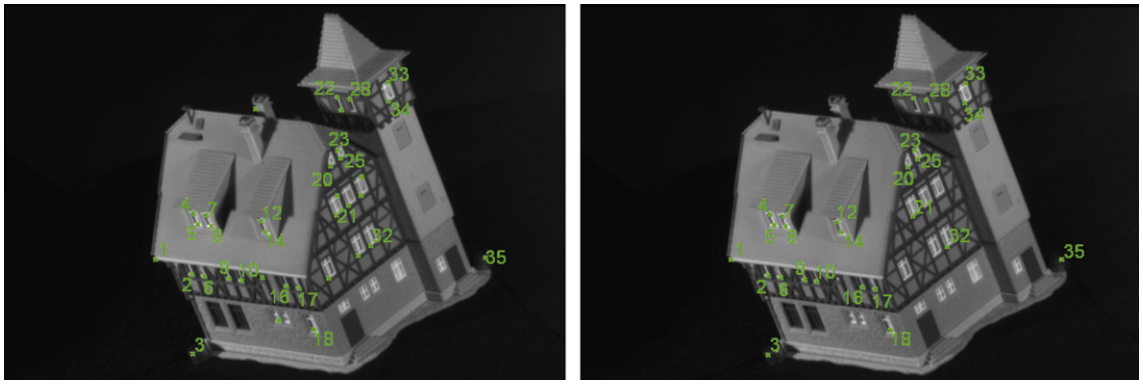


Fig. 7. Two successive images (frames 1 and 2) from the CMU House benchmark (http://vasc.ri.cmu.edu/idb/html/motion/house/index.html). The green labels are obtained from the matching procedure. The outcome is in this case perfect. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

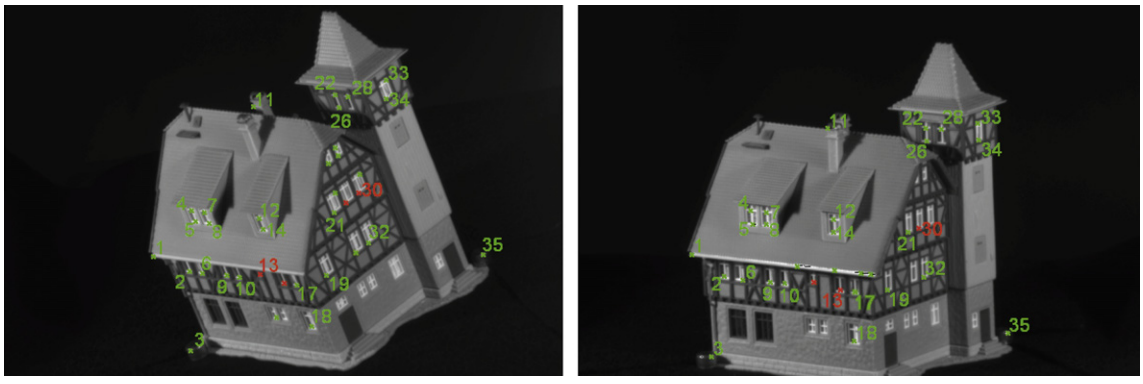


Fig. 8. Two random images (frames 1 and 67) considered for the matching procedure over the CMU House benchmark (http://vasc.ri.cmu.edu/idb/html/motion/house/index.html). The green labels correspond to correctly matched points (true matches and true singles), whereas the red labels mark wrong outcomes (false matches and false singles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Outcomes of the matching procedure tested on 109 pairs of feature sets extracted from 110 images in (http://vasc.ri.cmu.edu/idb/html/motion/house/index.html). The results are averages over the 109 tests. The top line refers to the 109 pairs of sequential images, whereas the bottom one to 109 pairs of randomly extracted images.

Input	Output performance			
Image pairs	True matches	True singles	False matches	False singles
Sequential	93.32%	5.48%	0.73%	0.47%
Random	75.86%	16.20%	5.68%	2.26%

formance against large transformations and the presence of feature occlusions. (Both studies focus on known issues reported for algorithms as in Scott and Longuet-Higgins (1991)). Table 4 displays the results as averages over the 109 tests. The outcomes of both studies appear to sensibly improve those in (Carcassoni and Hancock, 2003), and to remarkably improve those in (Wang and Hancock, 2006). Notice that the performance measure in (Carcassoni and Hancock, 2003; Wang and Hancock, 2006) is based exclusively on the second component of the pair of images, and hence slightly differs from the one used in this work, which we believe is more accurate. Also, the statistics in both (Carcassoni and Hancock, 2003; Wang and Hancock, 2006) are quite limited in sample size and image range.

4. Conclusions and future work

This article has proposed a versatile technique to perform point set matching over features extracted from images. The approach combines a number of ideas from related techniques in the literature. Its overall flexibility results from the possibility to define a library of metrics, from which similarity measures are selected and later employed over the specific matching problem. These heterogeneous measures are combined into a single pairing matrix, which is then manipulated via spectral techniques to obtain the actual matching.

The method has been tested on a number of different experimental studies, which have highlighted its performance, its robustness, as well as its computational scalability.

It is of future interest to come up with novel, descriptive metrics that can extend the applicability of the library and its usefulness in new domains of study.

References

Beauchemin, S.S., Barron, J.L., 1995. The computation of optical flow. ACM Comput. Surveys (CSUR) 27, 433–466.

Blondel, V.D., Gajardo, A., Heymans, M., Senellart, P., van Dooren, P., 2004. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. SIAM Rev. 46 (4), 647–666.

Carcassoni, M., Hancock, E.R., 2003. Correspondence matching with modal clusters. IEEE Trans. Pattern Anal. Machine Intell. 25 (12), 1609–1615.

Chui, H., Rangarajan, A., 2003. A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding 89 (2-3), 114–141.

Golub, G., Reinsch, C., 1970. Singular value decomposition and least squares solutions. Numerische Mathematik 14 (5), 403–420.

<http://vasc.ri.cmu.edu/idb/html/motion/house/index.html>.

Huttenlocher, D.P., Rucklidge, W.J. A multi-resolution technique for comparing images using the Hausdorff distance. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR 93), Jun 1993, pp. 705–706.

Jain, B., Vemuri, B.C. A robust algorithm for point set registration using mixture of gaussians. In: Proceedings of the International Conference on Computer Vision (ICCV 05), 2005, pp. 1246–1251.

Ma, D., Amonlirdviman, K., Raffard, R., Abate, A., Tomlin, C.J., Axelrod, J.D., 2008. Cell packing influences planar cell polarity signaling. Proc. National Academy Sci. 105 (48), 18800–18805.

Myronenko, A., Song, X., Carreira-Perpinan, M., 2007. Non-rigid point set registration: Coherent point drift. In: Schölkopf, B., Platt, J., Hoffman, T. (Eds.), Advances in Neural Information Processing Systems (NIPS) 19. MIT Press, Cambridge, MA, pp. 1009–1016.

Pilu, M. A direct method for stereo correspondence based on singular value decomposition. In: Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR 97), Jun 1997, pp. 261–266.

Sclaroff, S., Pentland, A.P., 1995. Modal matching for correspondence and recognition. IEEE Trans. Pattern Anal. Machine Intell. 17 (6), 545–561.

Scott, G.L., Longuet-Higgins, H.C., 1991. An algorithm for associating the features of two images. Proc. Royal Society London B-244, 21–26.

Shapiro, L.S., Brady, J.M., 1992. Feature-based correspondence: An eigenvector approach. Image Vision Comput. 10 (5), 283–288.

Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press, New York, NY, USA.

Silletti, A., Cenedese, A., Abate, A. The emergent structure of the Drosophila wing: a dynamic model generator. In: Proceedings of the International Conference on Computer Vision, Theory and Applications (VISAPP 09), Lisboa, PT, February 2009, pp. 406–410.

Srinivasan, S.H., Kankanhalli, M. Wide baseline spectral matching. In: Proceedings of the 2003 International Conference on Multimedia and Expo (ICME 03), Washington, DC, USA, 2003, pp. 93–96.

Torresani, L., Kolmogorov, V., Rother, C. Feature correspondence via graph matching: Models and global optimization. In: Proceedings of the European Conference on Computer Vision (ECCV 08), 2008, pp. 596–609.

Ullman, S., 1979. The Interpretation of Visual Motion. MIT Press, Cambridge, MA, USA.

Umeyama, S., 1988. An eigendecomposition approach to weighted graph matching problems. IEEE Trans. Pattern Anal. Machine Intell. 10 (5), 695–703.

Wang, H.F., Hancock, E.R., 2004. A kernel view of spectral point pattern matching. In: Structural, Syntactic, and Statistical Pattern Recognition. Lecture Notes in Computer Science, vol. 3138. Springer-Verlag, pp. 361–369.

Wang, H.F., Hancock, E.R., 2006. Correspondence matching using kernel principal components analysis and label consistency constraints. Pattern Recognit. 39 (6), 1012–1025.

Wells, W.M., 1997. Statistical approaches to feature-based object recognition. Internat. J. Comput. Vision 21 (1–2), 63–98.

Xiao, C.H., Nelson, H.C.Y., 2008. Corner detector based on global and local curvature properties. Optical Eng. 47 (5), 057008.