DaLI: Deformation and Light Invariant Descriptor

Edgar Simo-Serra · Carme Torras · Francesc Moreno-Noguer

Abstract Recent advances in 3D shape analysis and recognition have shown that heat diffusion theory can be effectively used to describe local features of deforming and scaling surfaces. In this paper, we show how this description can be used to characterize 2D image patches, and introduce DaLI, a novel feature point descriptor with high resilience to non-rigid image transformations and illumination changes. In order to build the descriptor, 2D image patches are initially treated as 3D surfaces. Patches are then described in terms of a heat kernel signature, which captures both local and global information, and shows a high degree of invariance to non-linear image warps. In addition, by further applying a logarithmic sampling and a Fourier transform, invariance to photometric changes is achieved. Finally, the descriptor is compacted by mapping it onto a low dimensional subspace computed using Principal Component Analysis, allowing for an efficient matching. A thorough experimental validation demonstrates that DaLI is significantly more discriminative and robust to illuminations changes and image transformations than state of the art descriptors, even those specifically designed to describe non-rigid deformations.

E. Simo-Serra · C. Torras · F. Moreno-Noguer (Corresponding author)
Institut de Robòtica i Informàtica Industrial
CSIC-UPC
08028, Barcelona, Spain
Tel.: +34-93-4015806
E-mail: {esimo,torras,fmoreno}@iri.upc.edu

1 Introduction

Building invariant feature point descriptors is a central topic in computer vision with a wide range of applications such as object recognition, image retrieval and 3D reconstruction. Over the last decade, great success has been achieved in designing descriptors invariant to certain types of geometric and photometric transformations. For instance, the SIFT descriptor [30] and many of its variants [3,21,31,32,46] have been proven to be robust to affine deformations of both spatial and intensity domains. In addition, affine deformations can effectively approximate, at least on a local scale, other image transformations including perspective and viewpoint changes. However, as shown in Fig. 1, this approximation is no longer valid for arbitrary deformations occurring when viewing an object that deforms non-rigidly.

In order to match points of interest under non-rigid image transformations, recent approaches propose optimizing complex objective functions that enforce global consistency in the spatial layout of all matches [12,13, 24,40,41,47]. Yet, none of these approaches explicitly builds a descriptor that goes beyond invariance to affine transformations. An interesting exception is [26], that proposes embedding the image in a 3D surface and using a Geodesic Intensity Histogram (GIH) as a feature point descriptor. However, while this approach is robust to non-rigid deformations, its performance drops under light changes. This is because a GIH considers deformations as one-to-one image mappings where image pixels only change their position but not the magnitude of their intensities.

To overcome the inherent limitation of using geodesic distances, we propose a novel descriptor based on the Heat Kernel Signature (HKS) recently introduced for
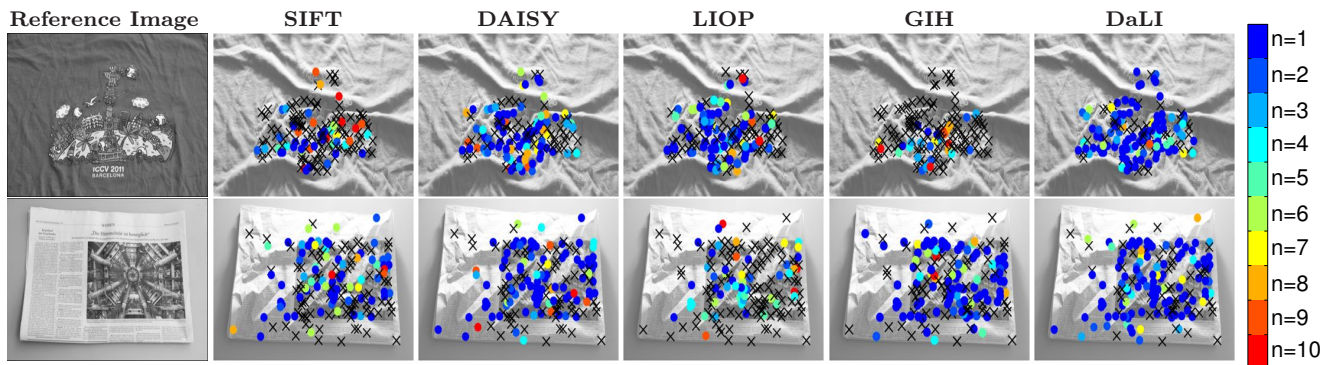
Fig. 1 Comparing DaLI against SIFT [30], DAISY [46], LIOP [53] and GIH [26]. Input images correspond to different appearances of the object shown in the reference images, under the effect of non-rigid deformations and severe changes of illumination. Colored circles indicate the match has been correctly found among the first n top candidates, where $n \leq 10$ is parameterized by the legend on the right. A feature is considered as mismatched when $n > 10$ and we indicate this with a cross. Note that the DaLI descriptor yields a significantly larger number of correct matches.

non-rigid 3D shape recognition [16, 39, 44], and which besides invariance to deformation, has been demonstrated to be robust to global isotropic [7] and even affine scalings [37]. In general, the HKS is particularly interesting in our context of images embedded on 3D surfaces, because illumination changes produce variations on the intensity dimension that can be seen as local anisotropic scalings, for which [7] still shows a good resilience.

Our main contribution is thus using the tools of diffusion geometry to build a descriptor for 2D image patches that is invariant to non-rigid deformations and photometric changes. To construct our descriptor we consider an image patch P surrounding a point of interest, as a surface in the $(x, y, \beta I(\mathbf{x}))$ space, where (x, y) are the spatial coordinates, $I(\mathbf{x})$ is the intensity value at (x, y), and β is a parameter which is set to a large value to favor anisotropic diffusion and retain the gradient magnitude information. Drawing inspiration from the HKS [16, 44], we then describe each patch in terms of the heat it dissipates onto its neighborhood over time. To increase robustness against 2D and intensity noise, we use multiple such descriptors in the neighborhood of a point, and weigh them by a Gaussian kernel. As shown in Fig. 1, the resulting descriptor (which we call DaLI, for Deformation and Light Invariant) outperforms state-of-the-art descriptors in matching points of interest between images that have undergone non-rigid deformations and photometric changes.

A preliminary version of this paper was already published in [33]. In the current work, we propose alternatives to both alleviate the high cost of the heat kernel computation and to reduce the dimensionality of the descriptor. More specifically, while in [33] the 3D embedding was performed considering a mesh with a uniform distribution of vertices in the (x, y) domain, here we investigate topologies with varying vertex densities.

This allows reducing the effective size of the underlying mesh, and hence to speed up the DaLI computation time by a factor of over 4. In addition, we have also compacted the size of the final descriptor by a factor of $50\times$ using a Principal Component Analysis (PCA) for dimensionality reduction. As a result, the descriptor we propose here can be computed and matched much faster when compared to [33], while preserving the discriminative power. For evaluation, we acquired a challenging dataset that contains 192 pairs of real images, manually annotated, of diverse materials under different degrees of deformation and being illuminated by very different illumination conditions. Fig. 1-left shows two samples of our dataset. We believe this is the first deformation and illumination dataset for evaluating image descriptors using real-world objects, and have made the dataset along with the code of the DaLI descriptor publicly available[1].

2 Related Work

The SIFT descriptor [30] has become the main reference among feature point descriptors, showing great success in capturing local affine deformations including scaling, rotation, viewpoint change and certain lighting changes. Since it is relatively slow to compute, most of the subsequent works have focused on developing faster descriptors [3, 10, 21, 31, 46]. Scale and rotation invariance has also been demonstrated in [22] using a combination of logarithmic sampling and multi-scale signal processing, although that requires large image patches which make the resulting descriptor more sensitive to other deformations. Indeed, as discussed in [52], little effort has
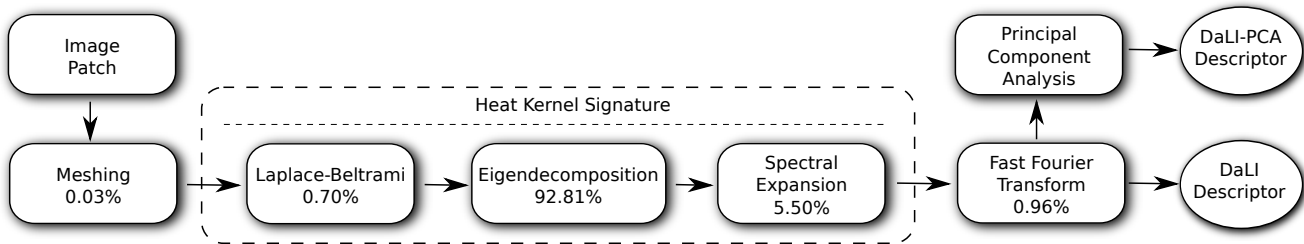
---

[1] http://www.iri.upc.edu/people/esimo/research/dali/

Fig. 2 Flowchart of the algorithm used to calculate the DaLI and DaLI-PCA descriptors. The percentages below each of the steps indicate the total amount of the contribution of that step to the computation time. Observe that 99% of the computation time corresponds to the Heat Kernel Signature calculation and specifically almost entirely to the eigendecomposition of the Laplace-Beltrami operator.

been devoted to building descriptors robust to more general deformations.

The limitations of the affine-invariant descriptors when solving correspondences between images of objects that have undergone non-rigid deformations are compensated by enforcing global consistency, both spatial and photometric, among all features [4,5,12,13, 24,40,41,47], or introducing segmentation information within the descriptor itself [48,49]. In any event, none of these methods specifically handles the non-rigid nature of the problem, and they rely on solving complex optimization functions for establishing matches.

An alternative approach is to directly build a deformation invariant descriptor. With that purpose, recent approaches in two-dimensional shape analysis have proposed using different types of intrinsic geometry. For example, [6,27] define metrics based on the inner-distance, and [28] proposes using geodesic distances. However, all these methods require the shapes to be segmented out from the background and represented by binary images, which is difficult to do in practice. In [26], it was shown that geodesic distances, in combination with an appropriate 3D embedding of the image, were adequate to achieve deformation invariance in intensity images. Nonetheless, this method assumes that pixels only change their image locations and not their intensities and, as shown in Fig. 1, is prone to failure under illumination changes.

There have also been efforts to build illumination invariant descriptors. Such works consider strategies based on intensity ordering and spatial sub-division [15,17,19, 18,20,45,53]. While these approaches are invariant to monotonically increasing intensity changes, their success rapidly falls when dealing with photometric artifacts produced by complex surface reflectances or strong shadows.

The DaLI descriptor we propose can simultaneously handle such relatively complex photometric and spatial warps. Following [26], we represent the images as 2D surfaces embedded in the 3D space. This is in fact

a common practice, although it has been mostly employed for low level vision tasks such as image denoising [43,56] or segmentation [55]. The fundamental difference between our approach and [26] is that we then describe each feature point on the embedded surface considering the heat diffusion over time [16,25,44] instead of using a Geodesic Intensity Histogram. As we will show in the results section this yields substantially improved robustness, especially to illumination changes. Heat diffusion theory has been used by several approaches for the analysis of 3D textured [23] and non-textured shapes [14,25,38,39], but to the best of our knowledge, it has not been used before to describe patches in intensity images.

One of the main limitations of the methods based on the heat diffusion theory is the high complexity cost they require. The bottleneck of their computation lies on an eigendecomposition of a $n_v \times n_v$ Laplacian matrix (see Fig. 3), where n_v is the number of vertices of the underlying mesh. This has been addressed by propagating the eigenvectors across different mesh resolutions [42,54] or using matrix exponential approximations [50]. In this paper, an annular multiresolution grid will be used to improve the efficiency of the DaLI computation. Additionally, PCA will be used to reduce the dimensionality of the original DaLI descriptor [33], hence speeding up the matching process as well.

3 Deformation and Light Invariant Descriptor

Our approach is inspired by current methods [16,44] that suggest using diffusion geometry for 3D shape recognition. In this section we show how this theory can be adapted to describe 2D local patches of images that undergo non-rigid deformations and photometric changes. A general overview of the different steps needed to compute the DaLI and DaLI-PCA descriptors can be seen in Fig. 2 and are explained more in detail below.
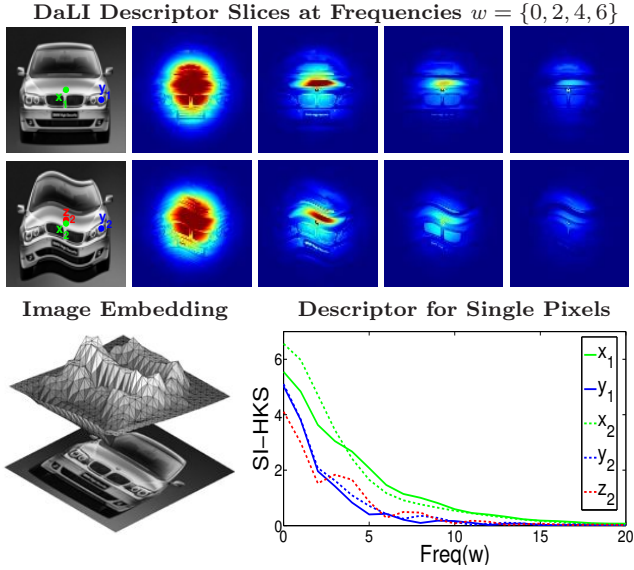
DaLI Descriptor Slices at Frequencies $w = \{0, 2, 4, 6\}$

Image Embedding   **Descriptor for Single Pixels**

Fig. 3 DaLI descriptor. Our central idea is to embed image patches in 3D surfaces and describe them based on heat diffusion processes. We represent the heat diffusion as a stack of images in the frequency domain. The top images show various slices of our descriptor for two different patches. The bottom-right graph depicts the value of the descriptor for the pixels marked by color circles in the upper images. Note that corresponding pixels have very similar signatures. However, the signature may significantly change from one pixel to its immediate neighbor. For instance, $\mathbf{z}_2$ is at one pixel distance from $\mathbf{x}_2$, but their signatures are rather different. As a consequence, using the signature of a single point as a descriptor is very sensitive to 2D noise in the feature detection process. We address this by simultaneously considering the signature of all the pixels within the patch, weighted by a Gaussian function of the distance to the center of the patch.

3.1 Invariance to Non-Rigid Deformations

Let us assume we want to describe a 2D image patch P, of size $S_P \times S_P$ and centered on a point of interest $\mathbf{p}$. In order to apply the diffusion geometry theory to intensity patches we regard them as 2D surfaces embedded in 3D space (Fig. 3 bottom-left). More formally, let $f : P \rightarrow M$ be the mapping of the patch P to a 3D Riemannian manifold M. We explicitly define this mapping by:

$$f : \mathbf{x} \rightarrow (x, y, \beta I(\mathbf{x})) \quad \forall \mathbf{x} \in P , \qquad (1)$$

where $I(\mathbf{x})$ is the pixel intensity at $\mathbf{x} = (x, y)^\top$, and β is a parameter that, as we will discuss later, controls the amount of gradient magnitude preserved in the descriptor.

Several recent methods [16, 25, 38, 39, 44] have used the heat diffusion geometry for capturing the local properties of 3D surfaces and performing shape recognition. Similarly, we describe each patch P based on the heat diffusion equation over the manifold M:

$$\left(\triangle_M + \frac{\partial}{\partial t} \right) u(\mathbf{x}, t) = 0 ,$$

where $\triangle_M$ is the *Laplace-Beltrami operator*, a generalization of the Laplacian to non-Euclidean spaces, and $u(\mathbf{x}, t)$ is the amount of heat on the surface point $\mathbf{x}$ at time t.

The solution $k(\mathbf{x}, \mathbf{y}, t)$ of the heat equation with an initial heat distribution $u_o(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{y})$ is called the *heat kernel*, and represents the amount of heat that is diffused between points $\mathbf{x}$ and $\mathbf{y}$ at time t, considering a unit heat source at $\mathbf{x}$ at time $t = 0$. For a compact manifold M, the heat kernel can be expressed by following spectral expansion [11, 38]:

$$k(\mathbf{x}, \mathbf{y}, t) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(\mathbf{x}) \phi_i(\mathbf{y}) , \qquad (2)$$

where $\{\lambda_i\}$ and $\{\phi_i\}$ are the eigenvalues and eigenfunctions of $\triangle_M$, and $\phi_i(\mathbf{x})$ is the value of the eigenfunction ϕ_i at the point $\mathbf{x}$. Based on this expansion, [44] proposes describing a point $\mathbf{p}$ on M using the Heat Kernel Signature

$$\mathrm{HKS}(\mathbf{p}, t) = k(\mathbf{p}, \mathbf{p}, t) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i^2(\mathbf{p}) , \qquad (3)$$

which is shown to be isometrically-invariant, and adequate for capturing both the local properties of the shape around $\mathbf{p}$ (when $t \rightarrow 0$) and the global structure of M (when $t \rightarrow \infty$).

However, while on smooth surfaces the HKS of neighboring points are expected to be very similar, when dealing with the wrinkled shapes that may result from embedding image patches, the heat kernel turns to be highly unstable along the spatial domain (Fig. 3 bottom-right). This makes the HKS particularly sensitive to noise in the 2D location of the keypoints. To handle this situation, we build the descriptor of a point $\mathbf{p}$ by concatenating the HKS of all points $\mathbf{x}$ within the patch P, properly weighted by a Gaussian function of the distance to the center of the patch. We therefore define the following Deformation Invariant (DI) descriptor:

$$\mathrm{DI}(\mathbf{p}, t) = [\mathrm{HKS}(\mathbf{x}, t) \cdot G(\mathbf{x}; \mathbf{p}, \sigma)]_{\forall \mathbf{x} \in P} , \qquad (4)$$

where $G(\mathbf{x}; \mathbf{p}, \sigma)$ is a 2D Gaussian function centered on $\mathbf{p}$ having a standard deviation σ, evaluated at $\mathbf{x}$. Note that for a specific time instance t, $\mathrm{DI}(\mathbf{p}, t)$ is a $S_P \times S_P$ array.

The price we pay for achieving robustness to 2D noise is an increase of the descriptor size. That is, if $\mathrm{HKS}(\mathbf{p}, t)$ is a function defined on the temporal domain
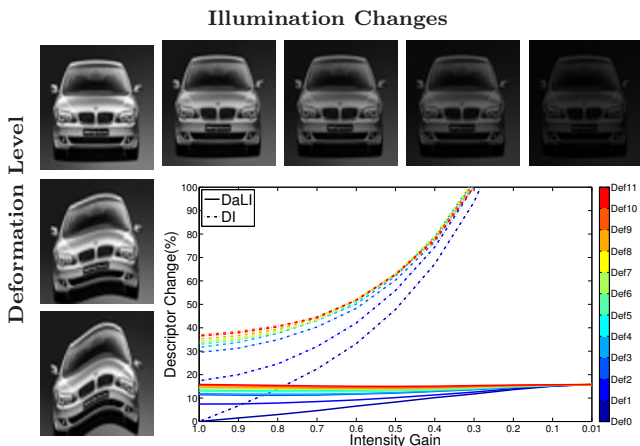
Fig. 4 Invariance of the DaLI and DI descriptors to non-rigid deformations and illumination changes. **Top row and left column images:** Different degrees of deformation and light changes applied on the top left reference patch P_0. Deformations are applied according to a function $\text{Def}(\cdot) \in \{\text{Def0}, \ldots, \text{Def11}\}$, where Def11 corresponds to the maximal deformation. Light changes are produced by scaling the intensity of P_0 by a gain $g \in [0,1]$. **Bottom Graph:** Given a deformation $\text{Def}(\cdot)$ and a gain factor g, we compute the percentage of change of the DI descriptor by $\|\text{DI}(P_0) - \text{DI}(\text{Def}(gP_0))\| / \|\text{DI}(P_0)\|$. The percentage of change for DaLI is computed in a similar way. Observe that DaLI is much less sensitive than DI, particularly to illumination changes.

$\mathbb{R}^+$ discretized into n_t equidistant intervals, the complete DI descriptor $\text{DI}(\mathbf{p}) = [\text{DI}(\mathbf{p}, t_1), \ldots, \text{DI}(\mathbf{p}, t_{n_t})]$ will be defined on $S_P \times S_P \times n_t$, the product of the spatial and temporal domains. However, note that for our purposes this is still feasible, because we do not need to compute a descriptor for every pixel of the image, but just for a few hundreds of points of interest. Furthermore, as we will next discuss, the descriptor may be highly compacted if we represent it in frequency domain instead of time domain and even further compacted by using dimensionality reduction techniques such as Principal Component Analysis (PCA).

3.2 Invariance to Illumination Changes

An inherent limitation of the descriptor introduced in Eq. (4) is that it is not illumination invariant. This is because light changes scale the manifold M along the intensity axis, and the HKS is sensitive to scaling. It can be shown that an isotropic scaling of the manifold M by a factor α, scales the eigenvectors and eigenvalues of Eq. (2) by factors $1/\alpha$ and $1/\alpha^2$, respectively [38]. The HKS of a point $\alpha\mathbf{p} \in \alpha M$ can then be written as

$$\text{HKS}(\alpha\mathbf{p}, t) = \sum_{i=0}^{\infty} e^{-\frac{\lambda_i}{\alpha^2}t} \frac{\phi_i^2(\mathbf{p})}{\alpha^2} = \frac{1}{\alpha^2}\text{HKS}(\mathbf{p}, \frac{t}{\alpha^2}), \quad (5)$$

which is an amplitude and time scaled version of the original HKS.

Nonetheless, under isotropic scalings, several alternatives have been proposed to remove the dependence of the HKS on the scale parameter α. For instance, [38] suggests normalizing the eigenvalues in Eq. (2). In this paper we followed [7], that applies three consecutive transformations on the HKS. First, the time-dimension is logarithmically sampled, which turns the time scaling into a time-shift, that is, the right-hand side of Eq. (5) begets $\alpha^{-2}\text{HKS}(\mathbf{p}, -2\log\alpha + \log t)$. Second, the amplitude scaling factor is removed by taking logarithm and derivative w.r.t. $\log t$. The Heat Kernel then becomes $\frac{\partial}{\partial \log t}\log\text{HKS}(\mathbf{p}, -2\log\alpha + \log t)$. The time-shift term $-2\log\alpha$ is finally removed using the magnitude of the Fourier transform, which yields SI-HKS$(\mathbf{p}, w)$, a scale invariant version of the original HKS in the frequency domain. In addition, since most of the signal information is concentrated in the low-frequency components, the size of the descriptor can be highly reduced compared to that of HKS$(\mathbf{p}, t)$ by eliminating the high-frequency components past a certain frequency threshold w_{max}.

As we will show in the results section, another advantage of the SI-HKS signature is that although it is specifically designed to remove the dependence of the HKS on isotropic scalings, it is quite resilient to anisotropic transformations, such as those produced by photometric changes that only affect the intensity dimension of the manifold M. Thus, we will use this signature to define our Deformation and Light Invariant (DaLI) descriptor:

$$\text{DaLI}(\mathbf{p}, w) = [\text{SI-HKS}(\mathbf{x}, w) \cdot G(\mathbf{x}; \mathbf{p}, \sigma)]_{\forall \mathbf{x} \in P} .$$

Again, the full DaLI$(\mathbf{p})$ descriptor is defined as a concatenation of w_{max} slices in the frequency domain, each of size $S_P \times S_P$.

Fig. 3-top shows several DaLI slices at different frequencies for a patch and a deformed version of it. As said above, observe that most of the signal is concentrated in the low frequency components. In Fig. 4 we compare the sensitivity of the DI and DaLI descriptors to deformation and light changes, simulated here by a uniform scaling of the intensity channel. Note that DaLI, in contrast to DI, remains almost invariant to light changes, and it also shows a better performance under deformations. In the results section, we will show that this invariance is also accompanied by a high discriminability, yielding significantly better results in keypoint matching than existing approaches.

In order to get deeper insight about the properties of the DaLI descriptor, we have further evaluated the HKS and SI-HKS descriptor variants on a synthetic experiment, in which we have rendered various sequences of images of a textured 3D wave-like mesh under different
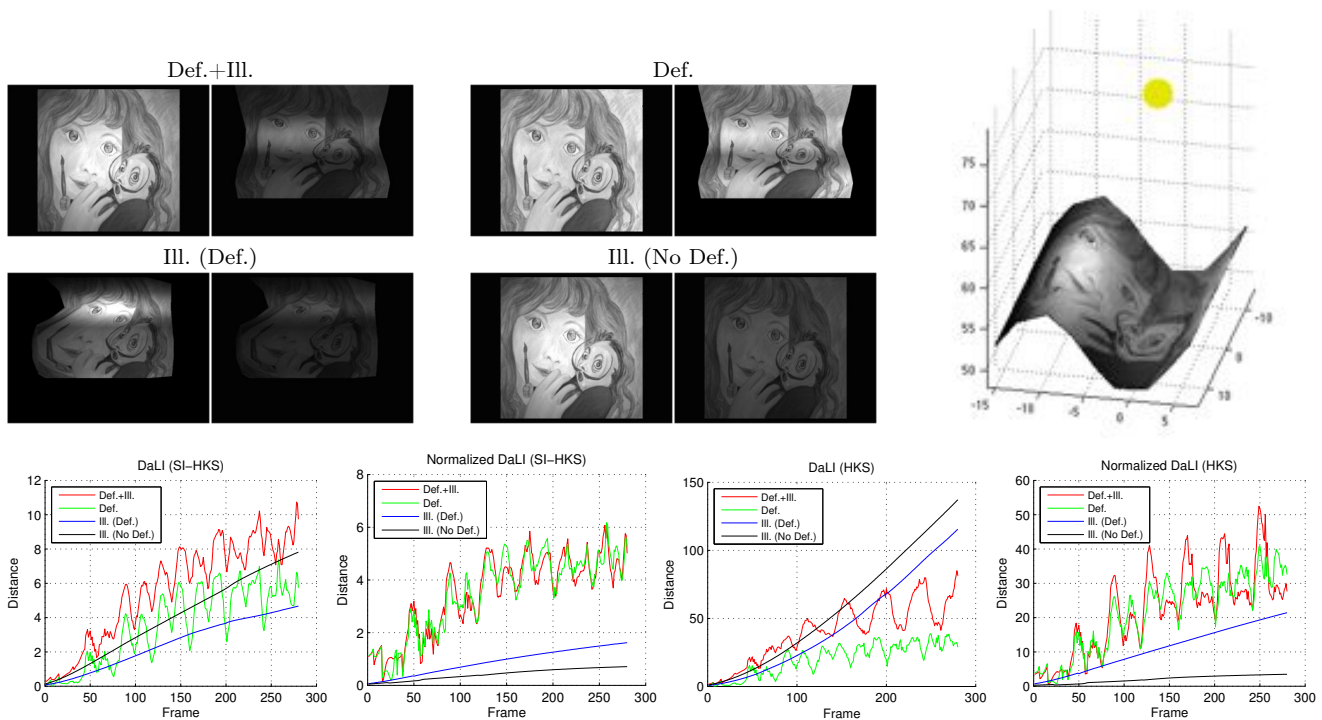
Fig. 5 Evaluation of the descriptor robustness on synthetic sequences. In the top-left we show two sample images (the reference image and one specific frame) from all four different scenarios we consider. In the top-right we show an 3D view of the rendering process, with the light position placed near the mesh and producing patterns of different brightness on top of the surface. The bottom row depicts the descriptor distance between every input frame and the reference image for different descriptor variants.

degrees of deformation and varying illumination conditions. The surface's reflectance is assumed to be Lambertian and the light source is moved near the surface, producing lighting patterns that combine both shading and the effects of the inverse-square falloff law.

We have analyzed four particular situations: *Def.+Ill.*, varying both deformation and the light source position; *Def.*, varying deformation and keeping the light source at infinity; *Ill. (Def.)*, starting with a largely deformed state which is kept constant along the sequence and varying the light source position; and *Ill. (No Def.)*, varying the light source position while keeping the surface flat. The mesh deformation in the first two sequences, corresponds to a sinusoidal warp, in which the amplitude of the deformation increases with the frame number. The varying lighting conditions in all experiments except the second, are produced by smoothly moving the light source on a hemisphere very close to the surface. Two frames from each of these sequences are shown in the top-left of Fig. 5.

For the evaluation, we computed the L2-norm distance between pairs of descriptors at the center of the first and n-th frames of the sequence. The results are depicted in Fig. 5-bottom. When computing the distances, we consider two situations: normalizing the in-

tensity of the input images so that the pixels follow a distribution $\mathcal{N}(0, 1)$, and directly using the input image intensities. The most interesting outcome of this experiment is to observe how the non-normalized SI-HKS descriptor has comparable distances for all the scenarios. On the other hand, the normalized versions (SI-HKS and HKS) seem to distinguish largely whether there is or there is not deformation. It is also worth noting that this normalization creates some instability at the earlier frames while the non-normalized SI-HKS descriptor starts at nearly 0 error and increases smoothly for all scenarios. Note also the low performance of the non-normalized HKS descriptor under illumination changes as seen by the exponential curves for the illumination scenarios *Ill. (Def)* and *Ill. (No Def)*, and the large fluctuations for both the deformation *Def* and the illumination changing scenario *Def. + Ill* . This indicates the importance of the logarithmic sampling and Fourier transform process we apply to make HKS illumination invariant.

3.3 Handling In-Plane Rotation

Although DaLI tolerates certain amounts of in-plane rotation, it is not designed for this purpose. This is be-
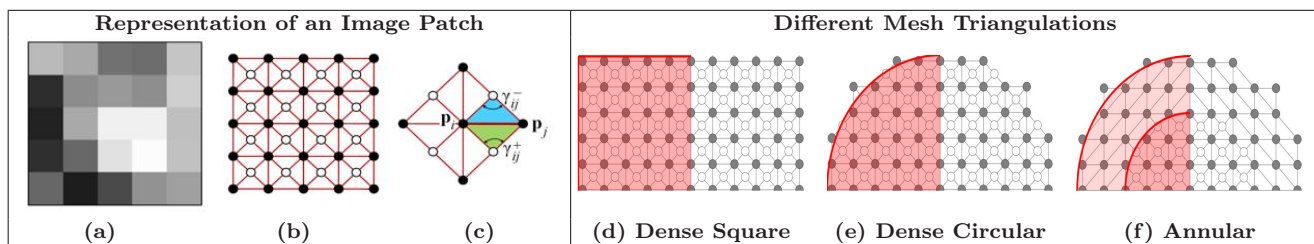
Fig. 6 Left: Patch representation. (a) Image patch. **(b)** Representation of the patch as a triangular mesh. For clarity of presentation we only depict the (x, y) dimension of the mesh. Note that besides the vertices placed on the center of the pixels (filled circles) we have introduced additional intra-pixel vertices (empty circles), that provide finer heat diffusion results and higher tolerance to in-plane rotations. **(c)** Definition of the angles used to compute the discrete Laplace-Beltrami operator. **Right: Several mesh triangulations.** Upper half of three different triangulations of a 11×11 image patch. The shading on the left half of the mesh indicates the density of the meshing. Dark red shading indicates high density and lighter red shading corresponds to low density. **(d)** Dense Square Mesh, with the same topology as in (b). By using circular meshes **(e, f)**, we reduce the number of vertices and thus, the computation time of the heat kernel. In the case of the annular mesh **(f)**, a further reduction of the number of nodes is achieved by having a variable resolution of the mesh that is more dense at the center. The edges of the annular mesh preserve symmetry around the central point in order to favor uniform heat diffusion.

cause with the aim of increasing robustness to 2D noise, we built the descriptor using all the pixels within the patch, and their spatial relations have been retained. Thus, if the patch is rotated, the descriptor will also be rotated.

In order to handle this situation, during the matching process we will consider several rotated copies of the descriptors. Therefore, given $\mathrm{DaLI}(\mathbf{p}_1)$ and $\mathrm{DaLI}(\mathbf{p}_2)$ we will compare them based on the following metric

$$\mathrm{d}(\mathbf{p}_1, \mathbf{p}_2) = \arg\min_{\theta_i} \|\mathrm{R}_{\theta_i}(\mathrm{DaLI}(\mathbf{p}_1)) - \mathrm{DaLI}(\mathbf{p}_2)\|$$

where $\| \cdot \|$ denotes the L_2 norm and $\mathrm{R}_{\theta_i}(\mathrm{DaLI}(\mathbf{p}))$ rotates $\mathrm{DaLI}(\mathbf{p})$ by an angle θ_i. This parameter is chosen among a discrete set of values $\boldsymbol{\theta}$.

This rotation handling will not be necessary when using Principal Component Analysis to compress the descriptor size as we describe in Section 5.2.

3.4 Implementation Details

We next describe a number of important details to be considered for the implementation of the DaLI descriptor.

3.4.1 Geometry of the embedding

For the numerical computation of the heat diffusion, it is necessary to discretize the surface. We therefore represent the manifold M on which the image patch is embedded using a triangulated mesh. Fig. 6(b) shows the underlying structured 8-neighbour representation we use. Although it requires introducing additional *virtual* vertices between the pixels, its symmetry with respect to the x and y directions provides robustness to

Table 1 DaLI computation time and mesh complexity for different triangulations of a circular patch with outer radius $S = 20$, and inner radius $S_o = 10$ (for the Annular mesh).

Mesh Type	# Pixels	# Vertices (n_v)	# Faces (n_f)	Time (s)
Dense Square	1681	3281	6400	1.988
Dense Circular	1345	2653	5144	1.509
Annular	1345	1661	3204	0.460

small amounts of rotation, and more uniform diffusions than other configurations.

As seen in Fig. 2, nearly all the computation time of the DaLI descriptor is spent calculating the Laplace-Beltrami eigenfunctions of the triangulated mesh. In the following subsection we will show that this computation turns to have a cubic cost on the number of vertices of the mesh, hence, important speed gains can be achieved by lowering this number. For this purpose we further considered a circular mesh (Fig. 6(e)), and a mesh with a variable density, like the one depicted in Fig. 6(f), where a lower resolution annulus is used for the pixels further away from the center.

By using an annular mesh with an inner radius $S_o = S/2$, where S is the size of the outer radius, we were able to speed up the computation of the DaLI descriptor by a factor of four compared to the Dense Squared configuration (see Table 1). Most importantly, this increase in speed did not result in poorer recognition rates.

Another important variable of our design is the magnitude of the parameter β in Eq. (1), that controls the importance of the intensity coordinate with respect to the (x, y) coordinates. In particular, as shown in Fig. 7, large values of β allow our descriptor to preserve edge information. This is a remarkable feature of the DaLI descriptor, because besides being deformation and il-
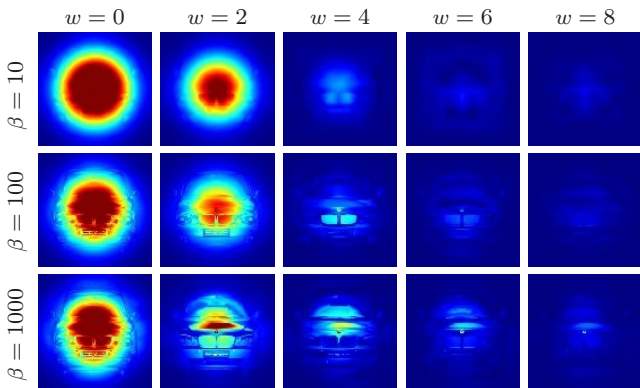
Fig. 7 Preserving edge information. Larger values of the parameter β in Eq. (1) allow the descriptor to retain edge information. Each row depicts the DaLI descriptor at frequencies $w = \{0, 2, 4, 6, 8\}$ for a different value of β computed on the car image from Fig. 3. Observe that for low values of β there is blurring on the higher frequencies of the descriptor.

Table 2 DaLI Parameters

Symbol	Parameter Description. (Default Value)
S	Outer radius of the annulus. (20)
S_o	Inner radius of the annulus. (10)
β	Magnitude of the embedding. (500)
σ	Standard deviation of Gaussian weighting. ($\frac{S}{2}$)
n_λ	# of eigenvectors of the Laplace-Beltrami operator. (100)
n_t	# of intervals in the temporal domain. (100)
w_{max}	# of frequency components used. (10)
θ_i	Rotation angles for descriptor comparison. ($\{-5, 0, +5\}$)
n_v	# of mesh vertices.(1661)
n_f	# of triangular faces in the mesh. (3204)
n_{pca}	# of PCA components for the DaLI-PCA. (256)

lumination invariant, edge information is useful to discriminate among different patches.

3.4.2 Discretization of the Laplace-Beltrami operator

In order to approximate the Laplace-Beltrami eigenfunctions on the triangular mesh we use the cotangent scheme described in [36]. We next detail the main steps.

Let $\{\mathbf{p}_1, \ldots, \mathbf{p}_{n_v}\}$ be the vertices of a triangular mesh, associated to an image patch embedded on a 3D manifold. We approximate the discrete Laplacian by a $n_v \times n_v$ matrix $\mathbf{L} = \mathbf{A}^{-1}\mathbf{M}$ where $\mathbf{A}$ is a diagonal matrix in which $\mathbf{A}_{ii}$ is proportional to the area of all triangles sharing the vertex $\mathbf{p}_i$. $\mathbf{M}$ is a $n_v \times n_v$ sparse matrix computed by:

$$\mathbf{M}_{ij} = \begin{cases} \sum_k m_{ik} & \text{if } i = j \\ -m_{ij} & \text{if } \mathbf{p}_i \text{ and } \mathbf{p}_j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

where $m_{ij} = \cot \gamma_{ij}^+ + \cot \gamma_{ij}^-$, and γ_{ij}^+ and γ_{ij}^- are the two opposite angles depicted in Fig. 6(c), and the subscript 'k' refers to all neighboring vertices of $\mathbf{p}_i$.

The eigenvectors and eigenvalues of the discrete Laplace-Beltrami operator can then be computed from the solution of the generalized eigenproblem $\mathbf{M}\boldsymbol{\Phi} = \boldsymbol{\Lambda}\mathbf{A}\boldsymbol{\Phi}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues $\{\lambda_i\}$ and the columns of $\boldsymbol{\Phi}$ correspond to the eigenvectors $\{\boldsymbol{\phi}_i\}$ in Eq. (2).

Note that the computational cost of the eigendecomposition is cubic in the size of $\mathbf{M}$, i.e., $\mathcal{O}(n_v^3)$. As discussed in the previous subsection, we mitigate this cost by choosing mesh topologies where the number of vertices is reduced. In addition, since the eigenvectors $\boldsymbol{\phi}_i$ with smallest eigenvalues have the most importance when calculating the HKS from Eq. (3), we can

approximate the actual value by only using a subset formed by the n_λ eigenvectors with smallest eigenvalues. Both these strategies allow the HKS calculation to be tractable in terms of memory and computation time.

Finally, Table 2 summarizes all the parameters that control the shape and size of the DaLI descriptor. The way we set their default values, shown between the parentheses, will be discussed in Section 5.1.

4 Deformation and Varying Illumination Dataset

In order to properly evaluate the deformation and illumination invariant properties of the DaLI descriptor and compare it against other state-of-the-art descriptors, we have collected and manually annotated a new dataset of deformable objects under varying illumination conditions. The dataset consists of twelve objects of different materials with four deformation levels and four illumination conditions each, for a total of 192 unique images. All images have a resolution of 640×480 pixels and are grayscale.

The types of objects in the dataset are 4 shirts, 4 newspapers, 2 bags, 1 pillowcase and 1 backpack. They were chosen in order to evaluate all methods against as many different types of deformation as possible. The objects can be seen in the top of Fig. 8.

4.1 Deformation and Illumination Conditions

The pipeline to acquire the images of each object consisted of, while keeping the deformation constant, changing the illumination before proceeding to the next deformation level. All images were taken in laboratory
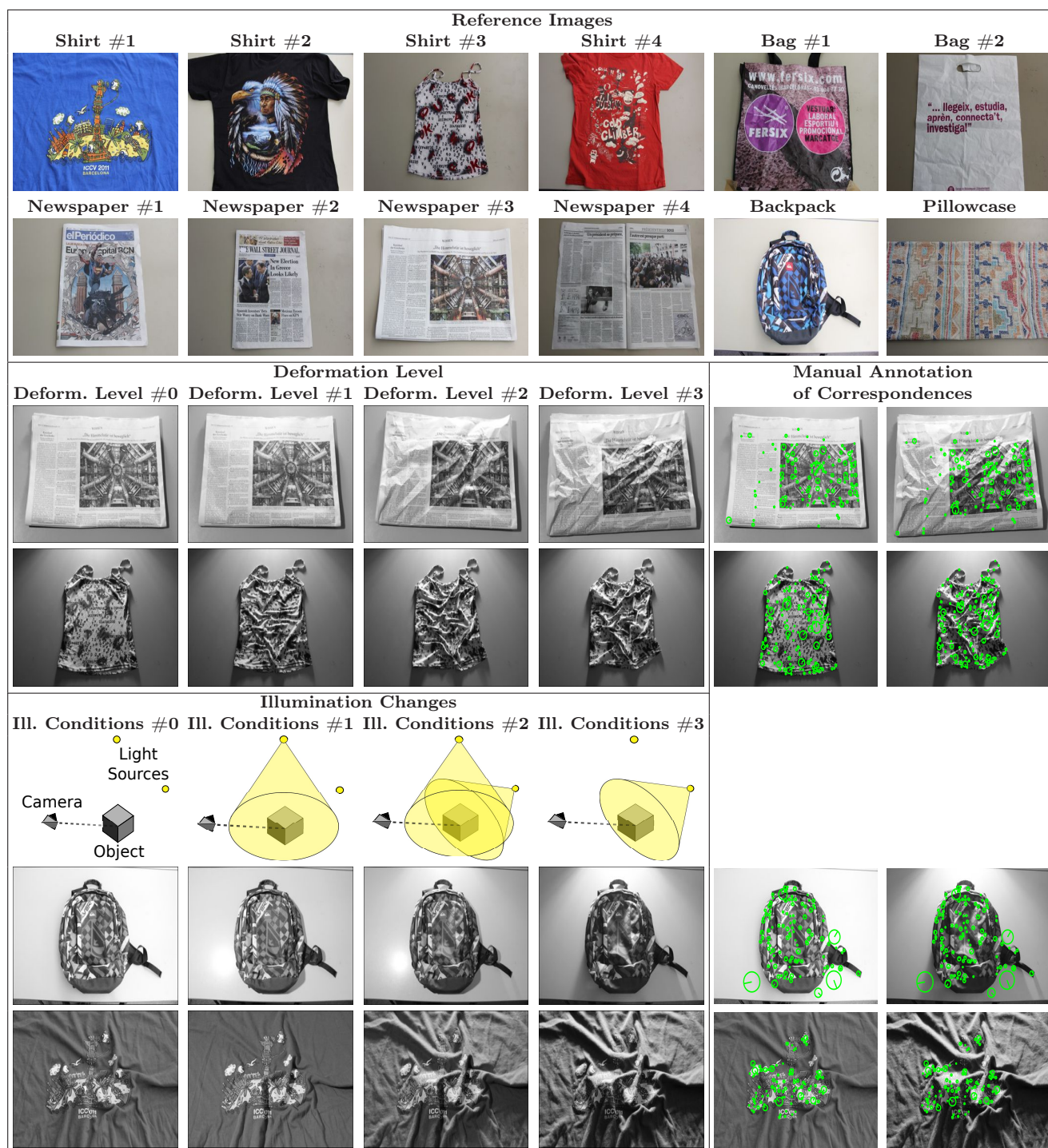
Fig. 8 Deformable and Varying Illumination dataset. **Top:** Reference images of the twelve objects in the dataset. Each object has four deformation levels and four illumination levels yielding a total of 16 unique images per object. **Middle-left:** Sample series of images with increasing deformation levels, and constant illumination. **Bottom-left:** Sample images of the different illumination conditions taken for a deformation level of each object. The illumination conditions #0, #1, #2 and #3 correspond to no illumination, global illumination, global+local illumination, and local illumination, respectively. **Middle-right and bottom-right:** Examples of feature points matched across image pairs. The first column corresponds to the reference image for the object. These feature points are detected using Differences of Gaussians (DoG) and are matched by manual annotation. Each feature point consists of image coordinates, scale coordinates and orientation.
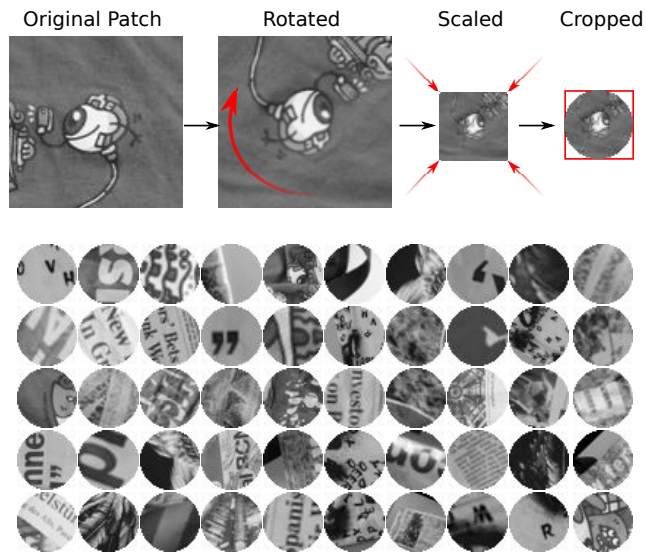
Fig. 9 Top: Outline of the process used to obtain patches for evaluating image descriptors. For each feature point, we initially extract a square patch centered on the feature, and whose size is proportional to the scale factor of the interest point. The patch is then rotated according to the orientation of the feature point, and finally scaled to a constant size and cropped to be in a circular shape. **Bottom:** Sample patches from the dataset, already rotated and scaled to a constant size in order to make them rotation and scale invariant.

conditions in order to fully control the settings for a suitable evaluation.

The reference image was acquired from an initial configuration where the object was straightened out as much as possible. While deformations are fairly subjective, as they were done incrementally over the previous deformation level, they are representative of increasing levels of deformation. Different deformation levels of an object with the same illumination conditions are shown in the middle-left of Fig. 8.

The illumination changes were produced by using two high power focus lamps. The first one was placed vertically over the object, at a sufficient distance to guarantee a uniform global illumination of the object's surface. The second lamp was placed at a small elevation angle and close to the object, in order to produce harsh shadows and local illumination artifacts. By alternating the states of these lamps, four different illumination levels are achieved: no illumination, global illumination, global with local illumination, and local illumination. The different illumination conditions for constant deformation levels can be seen in the bottom-left of Fig. 8. Note that even with moderate deformations, the presence of the local illumination causes severe appearance changes.

4.2 Manual Annotations

To build the ground truth annotations, we initially detected interest points in all images using a multi-scale Difference of Gaussians filter [30]. This yielded approximately between $500-600$ feature points per image, each consisting of a 2D image coordinate and its associated scale.

These feature points were then manually matched for each deformation level against the undeformed reference image, resulting in three pairs of matched feature points. All matches were done with top-light illumination conditions (Ill. Conditions #1, Fig. 8) to facilitate the annotation task and maximize the number of repeated features between each pair of images. The matching process yielded between 100 and 200 point correspondences for each pair of reference and deformed images. The same feature points are used for all illumination conditions for each deformation level. The middle-right images of Fig. 8 show a few samples of our annotation. Note that the matched points are generally not near the borders of the image to avoid having to clip when extracting image patches.

As we will discuss in the experimental section, in this paper we seek to compare the robustness of the DaLI and other descriptors to only deformation and light changes. Yet, although the objects in the dataset are not globally rotated, the deformations do produce local rotations. In order to compensate for this we use the SIFT descriptor as done in [31] to compute the orientation of each feature point, and align all corresponding features. When a feature point has more than one dominant orientation, we consider each of them to augment the set of correspondences.

4.3 Evaluation Criteria

In order to perform fair comparisons, we have developed a framework to evaluate local image descriptors on even grounds. This is done by converting each feature point into a small image patch which is then used to compute descriptors. This allows the evaluation of the exact same set of patches for different descriptors.

For each feature point we initially extract a square patch around it, with a size proportional to the feature point's scale. In the Experimental Section 5.3 we discuss the value of the proportionality constant we use. The patch is then rotated by the feature point's orientation using bilinear interpolation, and scaled to a constant size, which we have set to 41×41 pixels as done in [31]. Finally, the patch is cropped to a circular shape. This results in a scale and rotation invariant circular image

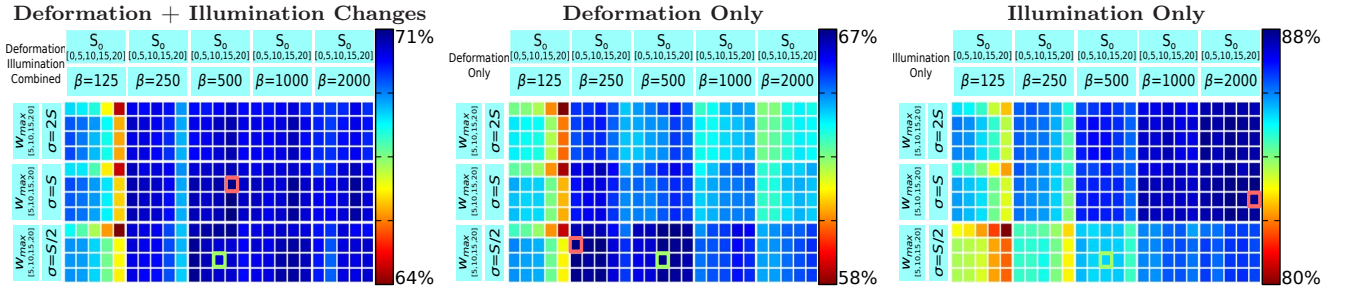**Fig. 10** DaLI performance for different values of the parameters S_o, β, σ and w_{max}. We compute and average the matching rate for the Shirt #1 and the Newspaper #1 objects in the dataset using $S_o \in \{0, 5, 10, 15, 20\}$ pixels, $\beta \in \{125, 250, 500, 1000, 2000\}$, $\sigma \in \{\frac{S}{2}, S, 2S\}$ and $w_{max} \in \{5, 10, 15, 20\}$ for three scenarios: both deformation and illumination changes, only deformation changes, and only illumination changes. The graphs depict the results of this 4D parameter exploration, where the color of each square represents the percentage of correctly matched points for a specific combination of the parameters. In order to visualize the differences, we scale the values separately for each scenario. The best parameters for each scenario are marked in red and can be seen to vary greatly amongst themselves. We use a compromise, and for all the experiments in this section we set these parameters (highlighted in green) to $\beta = 500$, $S_o = 10$, $\sigma = \frac{S}{2}$ and $w_{max} = 10$.

patch with a diameter of 41 pixels. The steps for extracting the patches are outlined in the top of Fig. 9, and the bottom of the figure shows a few examples of patches from the dataset.

Given these "normalized patches" we then assess the performance of the descriptors as follows. For each pair of reference/deformed images, we extract the descriptors of all feature points in both images. We then compute the L_2 distance between all descriptors from the reference and the deformed image. This gives a distance matrix, which is rectangular instead of square due to the creation of additional feature points when there are multiple dominant orientations. Patches that have different orientations but share the same location are treated as a unique patch. As evaluation metric we use a descriptor-independent detection rate, which is defined for the n top matches as:

$$\text{Detection Rate}(n) = \frac{100 \cdot N_c(n)}{N} \,, \qquad (6)$$

where $N_c(n)$ is the number of feature points from the reference image that have the correct match among the top n candidates in the deformed image, and N is the total number of feature points in the reference image.

For the experimental results we will discuss in the following section, we consider three different evaluation scenarios: deformation and illumination, only deformation, and only illumination. In the first case we compare all combinations of deformation and illumination with respect to the reference image which has no additional illumination (ill. conditions #0) and no deformation (deform. level #0). This represents a total of 15 comparisons for each object. In the second case we consider only varying levels of deformation for each illumination condition, which yields 12 different comparisons per object (three comparisons per illumination level).

When only considering illumination, each deformation level is compared to all illumination conditions. Again, this gives rise to 12 comparisons per object (three comparisons per deformation level).

5 Experimental Results

We next present the experimental results, in which we discuss the following main issues: an optimization of the descriptor parameters, a PCA-based strategy for compressing the descriptor representation, and the actual comparison of DaLI against other state-of-the-art descriptors, for matching points of interest in the proposed dataset. Finally, we analyze specific aspects such as the performance of all descriptors in terms of their size, the benefits of normalizing the intensity of input images, and a real application in which the descriptors are compared when matching points of interest in real sequences of a deforming cloth and a bending paper.

5.1 Choosing Descriptor's Parameters

We next study the influence and set the values of the DaLI parameters of Table 2. As the size S_P of the patch is fixed, causing the descriptor radius S to be also fixed, we will look at finding the appropriate value of other parameters, namely the magnitude β of the embedding, the degree σ of smoothing within the patch, the inner radius of the annulus S_o and the dimensionality w_{max} of the descriptor in the frequency domain. In order to find their optimal values, we used two objects in the dataset (Shirt #1 and Newspaper #1), and computed matching rates of their feature points for a wide range of values for each of these parameters.
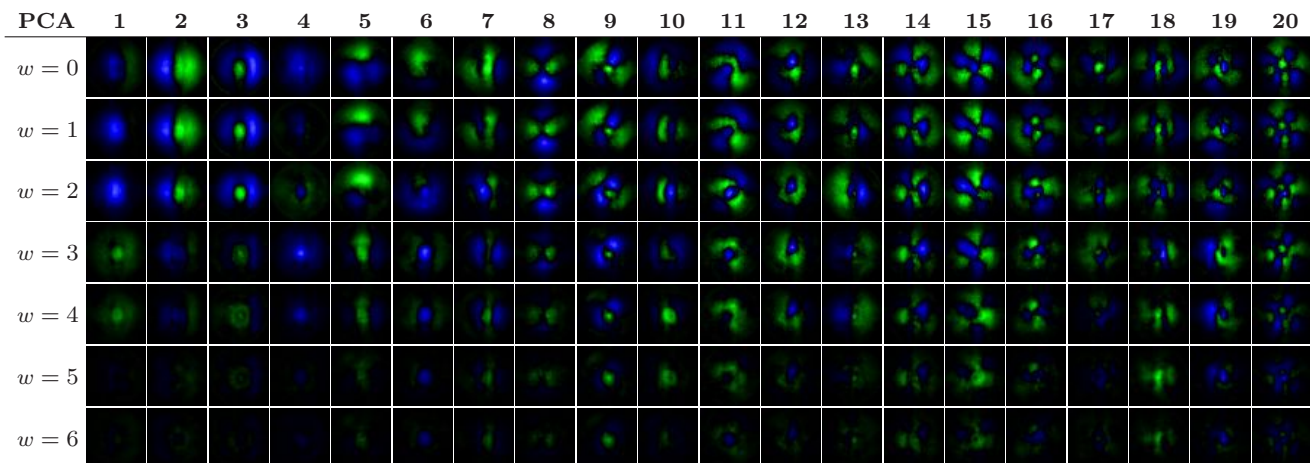
Fig. 11 The first 7 frequencies of the first 20 components of the PCA basis computed from images of two objects from the dataset (Shirt #1 and Newspaper #1). Each vector is normalized for visualization purposes. Positive values are displayed in green while negative values are displayed in blue. Most of the components do not contain much information at frequencies $w > 6$ and thus they are not displayed, although they are considered in the DaLI-PCA descriptor.

It is worth to point out that the number of eigenvectors n_λ of the Laplace-Beltrami operator was set to 100 in all cases. Note that this value represents a very small portion of all potential eigenvectors, in the order of two thousands (equal to the number of vertices n_v). Using a lesser number of them would eventually deteriorate the results, while not providing a significant gain in efficiency, and using more of them, almost did not improve the performance. Similarly, the number n_t of intervals in which the temporal domain is split is set to 100. Again, this parameter had almost no influence, neither in the performance of the descriptor nor in its computation time.

Fig. 10 depicts the results of the parameter sweeping experiment. We display the rates for three scenarios: when considering both deformation and illumination changes, only deformation changes, and only illumination changes. The most influential parameters are the weighting factor σ and to a lesser extent the magnitude of the embedding β. We see that for a wide range of parameters, the results obtained are very similar when considering both illumination and deformation, however, there is a balance to be struck between both deformation and illumination invariance. By increasing deformation invariance, illumination invariance is reduced and vice-versa. Finally we use a compromise, and the parameters we choose for all the rest of experiments are $\beta = 500$, $S_o = 10$, $\sigma = \frac{S}{2}$ and $w_{max} = 10$, besides the $n_\lambda = 100$ and $n_t = 100$ we mentioned earlier.

5.2 Compression with PCA

The DaLI descriptor has the downside of having a very high dimensionality, as its size is proportional to the
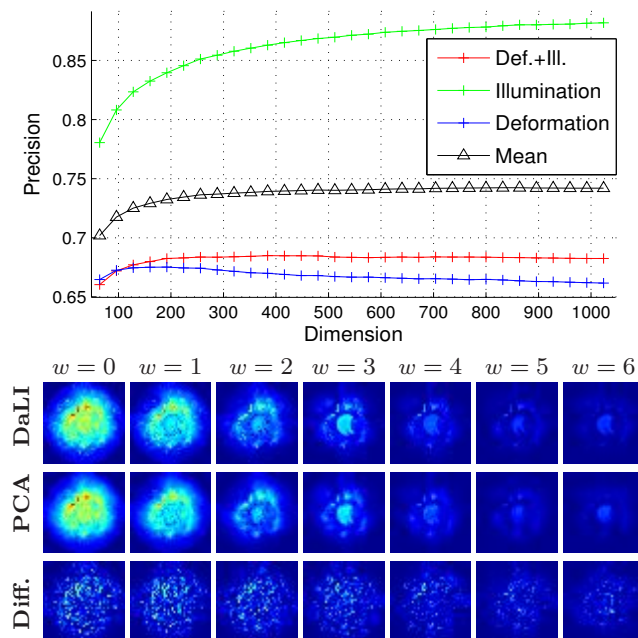

Fig. 12 Top: DaLI-PCA performance for different compression levels. Note that the overall mean precision does not vary much for $n_{pca} > 256$ components. **Bottom:** Comparison of an original DaLI descriptor with its compressed DaLI-PCA version obtained using 256 PCA components. For visualization purposes the values are normalized and the difference shown in the third row is scaled by $5\times$.

product of the number of vertices n_v used to represent the patch and the number of frequency components w_{max}. For instance, using patches with a diameter of 41 pixels and considering the first 10 frequency slices, results in a 13450-dimensional descriptor (1345 pixels by 10 frequency slices), requiring thus large amounts of memory and yielding slow comparisons. However, since
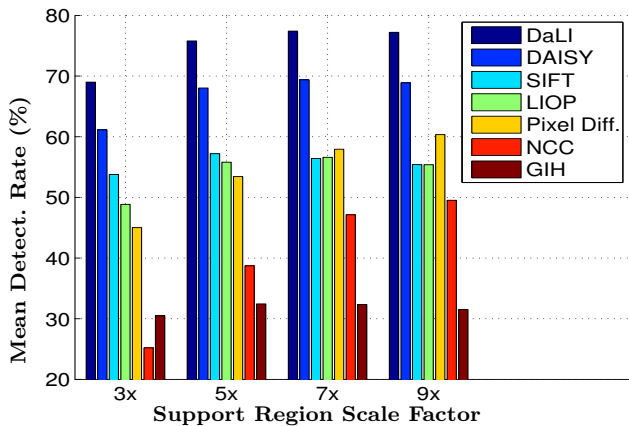
Fig. 13 Mean detection rates obtained by scaling regions of interest with different factors. While a 3× scale factor does lower the overall performance, the difference between a 5×, 7× or 9× scale factor is minimum for descriptors other than weighted pixel differences (Pix. Diff.) or normalized cross covariance (NCC), which do improve as interest regions increase in size. The results of the graph correspond to the average of the mean detection rates with Deformation+Illumination changes, Illumination-only changes and Deformation-only changes.

the descriptor is largely redundant, it can be compacted using dimensionality reduction techniques such as [8, 9, 35].

In this paper, as a simple proof of concept, we have used Principal Component Analysis for performing such compression. The PCA covariance matrix is estimated on 10436 DaLI descriptors extracted from images of the Shirt #1 and Newspaper #1. The $n_{pca} \ll n_v \cdot w_{max}$ largest eigenvectors are then used for compressing an incoming full-size DaLI descriptor. The resulting compacted descriptor, which we call DaLI-PCA, can be efficiently compared with other descriptors using the Euclidean distance. Fig. 11 shows the first 7 frequencies of the first 20 vectors of the PCA-basis. It is interesting to note that most of the information can be seen to be in the lower frequencies. This can be considered an experimental justification for the frequency cut off applied with the w_{max} parameter, which we have previously set to 10.

In order to choose the appropriate dimension n_{pca} of the PCA-basis, we have used our dataset to evaluate the matching rate of DaLI-PCA descriptors for different compression levels. The results are summarized in Fig. 12-top, and show that using fewer dimensions favors deformation invariance (actually, PCA can be understood as a smoothing that undoes some of the harm of deformations) while using more dimensions favors illumination invariance. The response to joint deformation and illumination changes does not improve after using between $200-300$ components, and this has been the criterion we used to set $n_{pca} = 256$ for the rest of

the experiments in this section. In Fig. 12-bottom we compare the frequency slices for an arbitrary DaLI descriptor and its approximation with 256 PCA-modes. Observe that the differences are almost negligible.

5.3 Comparison with Other Approaches

We compare the performance of our descriptors (both DaLI and DaLI-PCA) to that of SIFT [30], DAISY [46], LIOP [53], GIH [26], Normalized Cross Correlation (NCC) and Gaussian-weighted Pixel Difference. SIFT and DAISY are both descriptors based on Differences of Gaussians (DoG) and spatial binning which have been shown to be robust to affine deformations and to certain amount of illumination changes. LIOP is a recently proposed descriptor based on intensity ordering making it fully invariant to monotonic illumination changes. GIH is a descriptor specifically designed to handle non-rigid image deformations, but as pointed out previously, it assumes these deformations are the result of changing the position of the pixels within the image and not their intensity. NCC is a standard region-based metric known to possess illumination-invariant properties. Finally, we compare against a Gaussian-weighted pixel difference using the same convolution scheme as used for the DaLI descriptor. Standard parameters suggested in the original papers are used for all descriptors except for the LIOP descriptor in which using a larger number of neighboring sample points (8 instead of 4 neighbors) results in a higher performance at the cost of a larger descriptor (241920 instead of 144 dimensions). The LIOP and SIFT implementations are provided by VLfeat [51]. We use the authors' implementation of DAISY and GIH.

The evaluation is done on the dataset presented in Section 4. All the descriptors are therefore tested on exactly the same image patches in order to exclusively judge the capacity of local feature representation. Yet, as mentioned in Sec. 4.3, the dataset still requires setting the scale factor to use for the points of interest. This value corresponds to the relative size of each image patch with respect to the scale value obtained from the DoG feature point detector. For this purpose, we evaluated the response of all descriptors for scale factors of 3×, 5×, 7× and 9×. The results are shown in Fig. 13. Although the SIFT implementation uses a default value of 3×, we have observed that the performance of all descriptors improves by increasing the patch size. Note that this does not result in a higher computational cost, as the final size of the patch is normalized to a circular shape with a diameter of 41 pixels. The maximum global response for all descriptors is achieved when us-
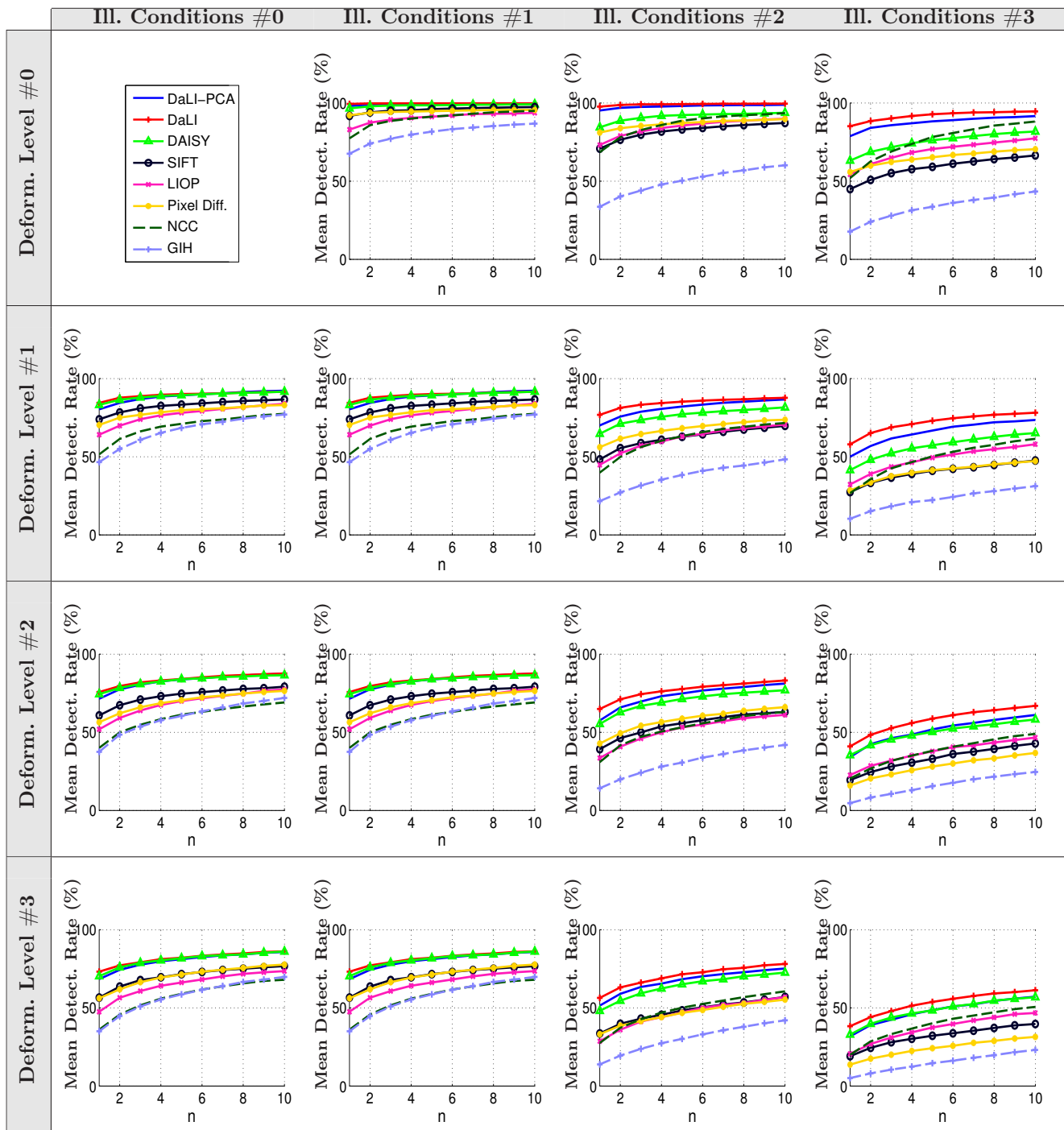
Fig. 14 Detection rate when simultaneously varying deformation level and illumination conditions. Each graph represents the average of the mean detection rate between the reference image (ill. conditions #0 and deform. level #0) and all images in the dataset under specific light and deformation conditions.

ing a 7× scale factor, which is the value we use for all the experiments reported below.

The results for concurrent deformation and illumination are summarized in Fig. 14. DaLI consistently outperforms all other descriptors, although the more favorable results are obtained under large illumination

changes. The performance of DAISY is very similar to that of DaLI when images are not affected by illumination artifacts. In this situation, the detection rates of DAISY are approximately between $2-5\%$ below to those obtained by DaLI. However, when illumination artifacts become more severe, the performance of
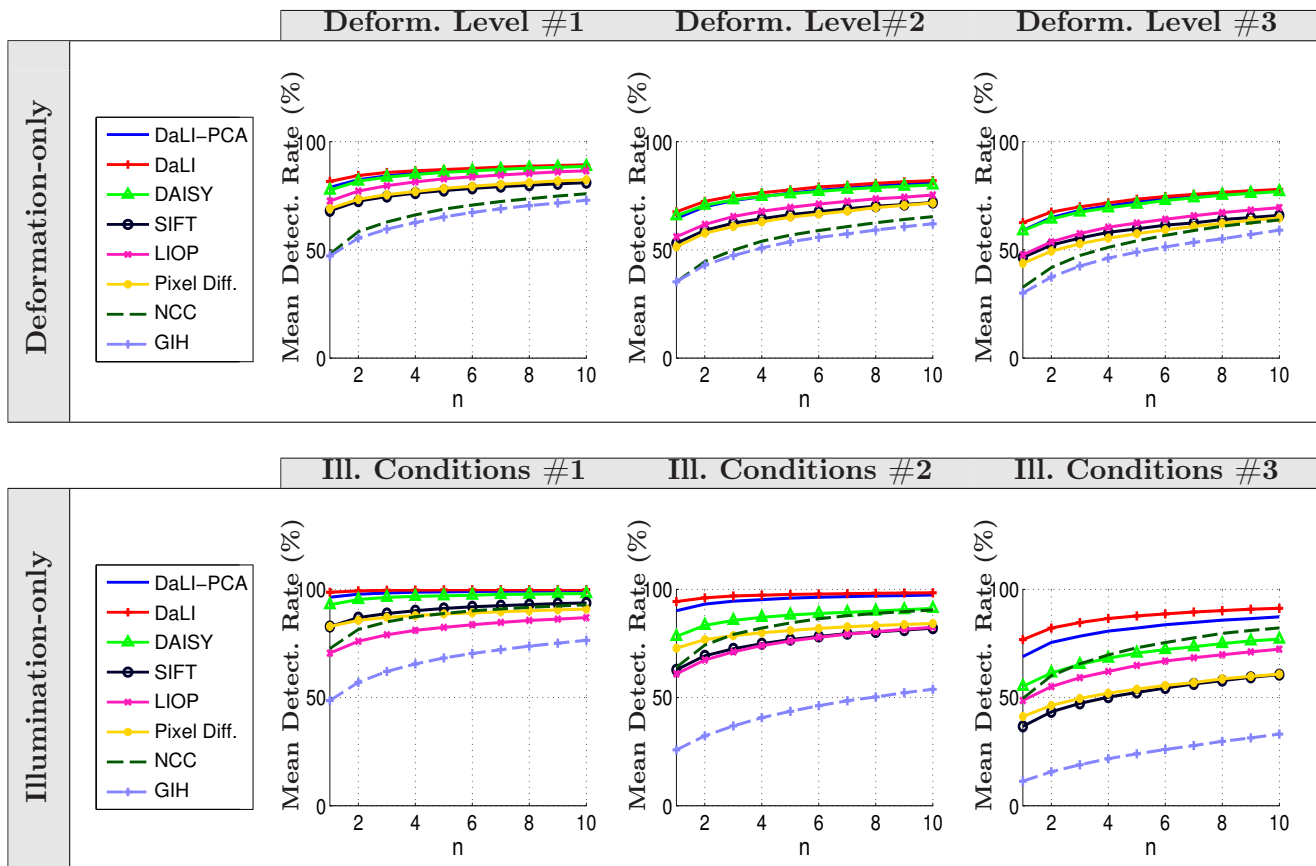
Fig. 15 Top: Results when varying only the deformation while keeping the illumination conditions constant. It can be seen that both DaLI and DAISY largely outperform the rest of descriptors. **Bottom:** Results of varying only the illumination conditions while keeping the deformation level constant. Note that only DaLI remains robust to illumination changes. The performance of DAISY falls roughly a 20% compared to DaLI.

DAISY rapidly drops, yielding detection rates which are more than 20% below DaLI. SIFT, LIOP, and Pixel Difference yield similar results, with SIFT being better at weak illumination changes and LIOP better at handling strong illumination changes. Yet, these three methods are one step behind DaLI and DAISY. NCC generally performs worse except in situations with large illumination changes, where it even outperforms DAISY. On the other hand, GIH performs quite poorly even when no light changes are considered. This reveals another limitation of this approach, in that it assumes the effect of deformations is to locally change the position of image pixels, while in real deformations some of the pixels may disappear due to occlusions. Although our approach does not explicitly address occlusions, we can partially handle them by weighing the contribution of the pixels within each patch, by a function decreasing with the distance to the center. Thus, most of the information of our descriptor is concentrated in a small region surrounding the point of interest, hence making it less sensitive to occlusions. The results also show that

the compressed DaLI-PCA follows a similar pattern as DaLI, and specially outperforms DAISY under severe illumination conditions.

In Fig. 15 we give stronger support to our arguments by independently evaluating deformations and illumination changes. These graphs confirm that under deformation-only changes, DaLI outperforms DaLI-PCA and DAISY by a small margin of roughly 3%. Next, SIFT, LIOP, and Pixel Difference yield similar results, roughly 20% below DaLI in absolute terms. GIH and NCC yield also similar results, although their performance is generally very poor. When only illumination changes are considered, both DaLI and DaLI-PCA significantly outperform other descriptors, by a margin larger than 20% when dealing with complex illumination artifacts. The only notable difference in this scenario is that the NCC descriptor outperforms SIFT and Pixel Difference. As GIH is not invariant to illumination changes, it obtains poor results. Similarly, since LIOP is designed to be invariant to monotonic lighting
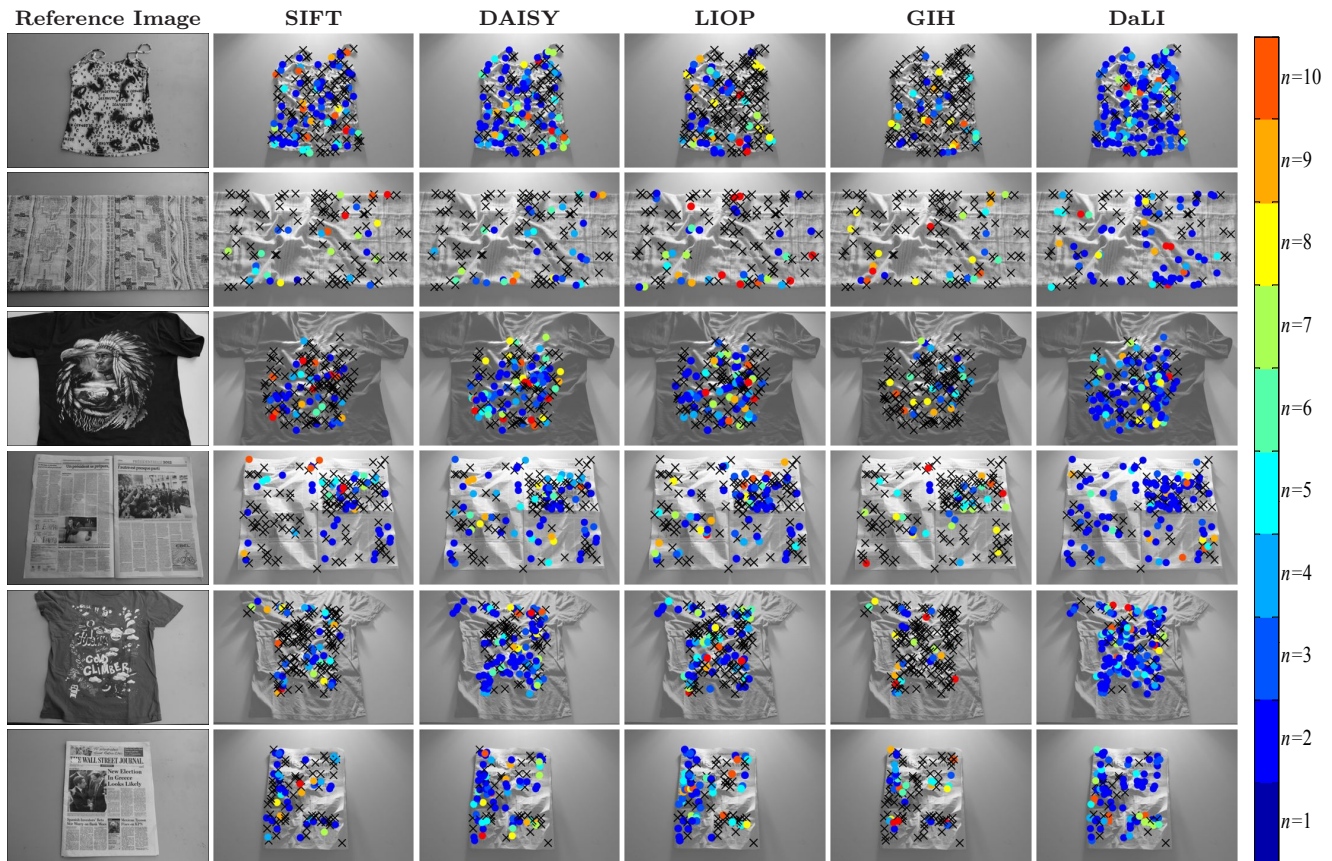
Fig. 16 Sample results from the dataset. As in Fig. 1, the color of the circles indicates the position n of the correct match among the top candidates. If $n > 10$ we consider the point as unmatched and mark it with a cross.

Table 3 Evaluation results on the dataset for all descriptors. Results are obtained by averaging the first match percentage values over all images being tested under all different conditions.

Descriptor	Deformation	Illumination	Deformation+Illumination
DaLI-PCA	67.425	85.122	68.368
DaLI	**70.577**	**89.895**	**72.912**
DAISY	67.373	75.402	66.197
SIFT	55.822	60.760	53.431
LIOP	58.763	60.014	52.176
Pixel Diff.	54.714	65.610	54.382
NCC	38.643	62.042	41.998
GIH	37.459	28.556	31.230

Fig. 17 Some of the true positive, false positive, false negative and true negative image patch pairs obtained using the DaLI descriptor on the dataset. Note that most of the false negatives are due to large orientation changes across feature points.

changes, it does not perform that well in real images that undergo complex illumination artifacts.

In summary, the experiments have shown that DaLI globally obtains the best performance. Its best relative response when compared with other descriptors is obtained when the deformations are mild and the light changes drastic. Some sample results on particular images taken from the dataset can be seen in Fig. 16. Ad-
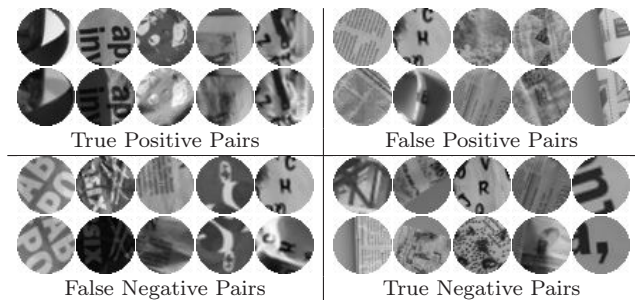
ditionally, numeric results for the best candidate ($n = 1$ in Eq. 6) under different conditions for all descriptors are shown in Table 3.

Finally, examples of particular patch matches are depicted in Fig. 17. The true positives pairs can be seen to be matched despite large changes. On the other hand, the false negatives seem largely generated by differences in orientations of the feature points: they correspond to the same patch, only rotated. The false positives share
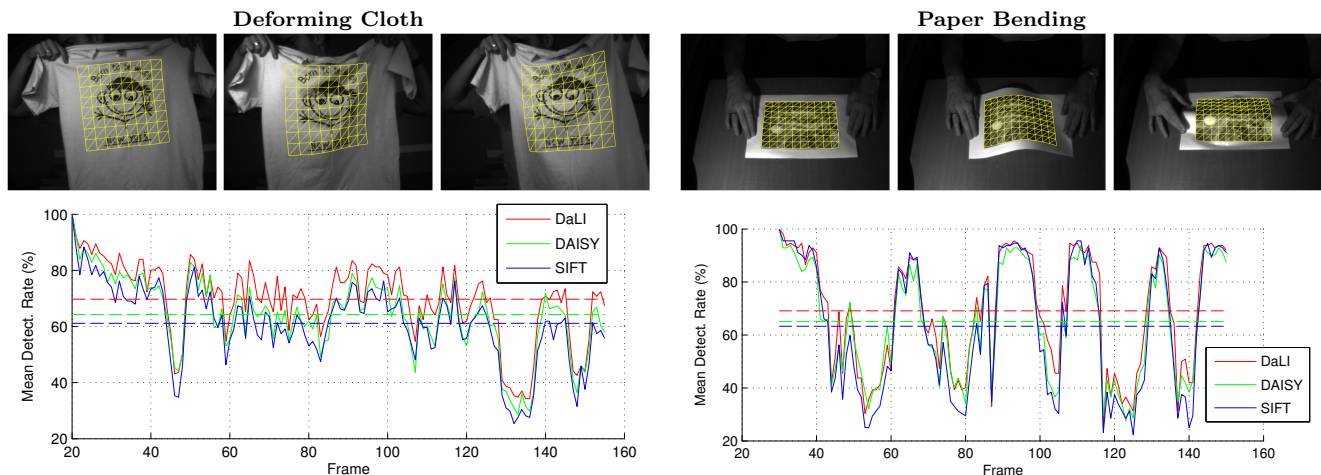
Fig. 18 Mean detection accuracy on two real world videos from [34]. In the top row we show three example frames from each video. In the bottom row we plot the accuracy for each frame for three descriptors: DaLI, DAISY and SIFT. Additionally the mean for each descriptor is displayed as a dashed line.

Table 4 Comparison of performance and descriptor size.

Descriptor	Size	Deform.	Illum.	Deformation+ Illumination
DaLI-PCA	128	**67.45**	82.34	67.71
SIFT	128	55.82	60.76	53.43
LIOP	144	54.01	44.89	44.45
DAISY	200	67.37	75.40	66.20
GIH	176	37.46	28.56	31.23
DaLI-PCA	**256**	67.43	**85.12**	**68.37**

some similarity, although they are mainly from heavily deformed images.

5.4 Descriptor Size Performance

Since larger descriptors may a priori have an unfair advantage, we next provide results of an additional experiment in which we compare descriptors having similar sizes. The LIOP we calculate in this case uses 4 neighbours instead of the 8 neighbours we considered before, which results in a smaller size, although also in a lower performance. GIH is originally 176-dimensional, thus the results are the same as in Table 3. NCC and Pixel Diff, are not considered for this experiment as their size is $41 \times 41 = 1681$.

Results are shown in Table 4. We can see that the 128-dimensional DaLI-PCA outperforms all other descriptors except the 256-dimensional DaLI-PCA. It is worth noting the large performance gain obtained over the standard SIFT descriptor.

5.5 Benefits of Intensity Normalization

We next extend the analysis we introduced in Sect. 3.2 in which we evaluated SI-HKS and HKS with and without pre-normalizing the intensity of input images. We will also consider SIFT and DAISY, which have been the most competitive descriptors in previous experiments. Since SIFT/DAISY implementations require the pixels to be in a $[0, 1]$ range, we have normalized each image patch so that the pixels follow the distribution $\mathcal{N}(0.5, (2 \cdot 1.956)^{-1})$. This makes it so that on average 95% of the pixels will fall in $[0, 1]$. Pixels outside of this range are set to either 0 or 1.

We compare the DaLI descriptor (both its SI-HKS and HKS variants), DAISY and SIFT, with and without normalization. Results are shown in Table 5. We can see that for DAISY and SIFT, since they perform a final normalization stage, the results do not have any significant change. In the case of the DaLI descriptor, though, we see that there is a rather significant performance increase when using the SI-HKS variant over the HKS one, even with patch normalization. This demonstrates again that the role of the Fourier Transforms applied in HKS to make it illumination invariant go far beyond a simple normalization. In addition, SI-HKS compresses the descriptor in the frequency domain and is one order of magnitude smaller than the HKS variant.

5.6 Evaluation on Real World Sequences

This section describes additional experiments on two real world sequences of deforming objects, taken from [34]. One consists of a T-Shirt being waved in front of a camera (Deforming Cloth) and the other consists of a

Table 5 Effect of normalizing image patches for various descriptors.

Descriptor	Normalization?	Deformation	Illumination	Deformation+Illumination
DaLI (SI-HKS)	**No**	**70.58**	**89.90**	**72.91**
DaLI (SI-HKS)	Yes	70.38	88.60	72.28
DaLI (HKS)	No	66.27	84.21	67.83
DaLI (HKS)	Yes	67.20	84.62	69.42
DAISY	No	67.37	75.40	66.20
DAISY	Yes	67.08	75.59	66.27
SIFT	No	55.82	60.76	53.43
SIFT	Yes	55.05	61.83	53.21

piece of paper being bended in front of a camera (Paper Bending). We use points of interest computed with the Differences of Gaussians detector (DoG) and follow the same patch extraction approach as in the rest of the paper. The points of interest are calculated for the first frame in each sequence and then propagated using the provided 3D ground truth to the other frames. We use the same descriptor parameters as in the rest of the experiments, and seek to independently match the points of interest in the first frame to those of all the other frames.

As we can observe in Fig. 18, DaLI outperforms both DAISY and SIFT[2]. We obtain a 5.5% improvement over DAISY on the Deforming Cloth sequence and a 4.1% improvement on the Paper Bending sequence. Note that these sequences do not have as complicated illumination artifacts as our dataset, an unfavorable situation for our descriptor. Yet, DaLI still consistently outperforms other approaches along the whole sequence.

6 Discussion and Conclusions

Heat diffusion theory has been recently shown effective for 3D shape recognition tasks. In this paper, we have proposed using these tools to build DaLI, a feature point descriptor for 2D image patches, that is very robust to both non-rigid deformations and illumination changes. The advantages of our method with respect to the state-of-the-art have been demonstrated by extensively testing them on a new deformation and varying illumination evaluation dataset[3].

We have also shown that simple dimensionality reduction techniques such as PCA can be effectively used to reduce dimensionality while maintaining similar performance. This seems to give the intuition that further improvements can be obtained by using more advanced and powerful techniques such as LDAHash [8]. Work

has also been done in optimizing the calculation speed by means of more complex meshing to reduce the cost of computing the eigenvectors of the Laplace-Beltrami operator.

As part of future work we will investigate recent and promising alternatives to the Heat Kernel Signatures (HKS), such as the Wave Kernel Signature (WKS) [2], and strategies to directly learn spectral descriptors in a supervised manner [1,29]. Using labeled training data would likely further increase the performance of our descriptor.

Additionally we will intend to make DaLI invariant to scale and rotations without the need to explore a wide range of discrete values. We will investigate two alternatives for this purpose: 1) incorporating prior information of the orientation and scale within each frequency slice, as it is done for the SIFT descriptor; 2) using a logarithmic sampling and Fourier Transform Modulus (FTM) as in [22].

Finally, we also plan to look into the function to weight the pixels within each patch. We are currently using a Gaussian distribution centered on the patch. However, there have been recent alternatives that compute similar functions based on segmentation information, which have shown to significantly improve the performance of standard descriptors [48,49].

7 Acknowledgements

References

---

[2] Again, we only compare against DAISY and SIFT, as these are the descriptors which have been more competitive in the experiments with the full dataset.

[3] http://www.iri.upc.edu/people/esimo/research/dali/

1. Yonathan Aflalo, Er M. Bronstein, Michael M. Bronstein, and Ron Kimmel. Deformable shape retrieval by learning diffusion kernels. In *International Conference on Scale Space and Variational Methods in Computer Vision*, 2011.

2. M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *Computer Vision Workshops (ICCV Workshops), IEEE International Conference on*, pages 1626–1633, Nov 2011.

3. H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *European Conference on Computer Vision*, pages 404–417, 2006.

4. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

5. A. Berg, T. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 26–33, 2005.

6. A. Bronstein, M. Bronstein, A. Bruckstein, and R. Kimmel. Analysis of two-dimensional non-rigid shapes. *International Journal of Computer Vision*, 78(1):67–88, 2007.

7. M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1704–1711, 2010.

8. M. Bronstein C. Strecha, A. Bronstein and Pascal Fua. LDA-Hash: Improved matching with smaller descriptors. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 34(1), 2012.

9. H. Cai, K. Mikolajczyk, and J. Matas. Learning linear discriminant projections for dimensionality reduction of image descriptors. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 33(2):338–352, 2011.

10. M. Calonder, V. Lepetit, M. Ozuysa, T. Trzcinski, C. Strecha, and P. Fua. BRIEF: Computing a local binary descriptor very fast. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 34(7):1281–1298, 2012.

11. I. Chavel. *Eigenvalues in Riemannian Geometry*. London Academic Press, 1984.

12. H. Cheng, Z. Liu, N. Zheng, and J. Yang. A deformable local image descriptor. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

13. M. Cho, J. Lee, and K. Lee. Feature correspondence and deformable object matching via agglomerative correspondence clustering. In *International Conference on Computer Vision*, pages 1280–1287, 2009.

14. Fernando de Goes, Siome Goldenstein, and Luiz Velho. A hierarchical segmentation of articulated bodies. In *Proceedings of the Symposium on Geometry Processing*, pages 1349–1356, 2008.

15. Bin Fan, Fuchao Wu, and Zhanyi Hu. Rotationally invariant descriptors using intensity order pooling. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 34(10):2031–2045, Oct 2012.

16. K. Gębal, J. A. Bærentzen, H. Aanæs, and R. Larsen. Shape analysis using the auto diffusion function. In *Proceedings of the Symposium on Geometry Processing*, pages 1405–1413, 2009.

17. R. Gupta and A. Mittal. Illumination and Affine-Invariant Point Matching using an Ordinal Approach. In *International Conference on Computer Vision*, 2007.

18. R. Gupta, H. Patil, and A. Mittal. Robust order-based methods for feature description. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

19. Raj Gupta and Anurag Mittal. Smd: A locally stable monotonic change invariant feature descriptor. In *European Conference on Computer Vision*, pages 265–277, 2008.

20. Marko Heikkilä, Matti Pietikäinen, and Cordelia Schmid. Description of interest regions with local binary patterns. *Pattern Recognition*, 42(3):425–436, March 2009.

21. Y. Ke and R. Sukthankar. PCA-SIFT: a more distinctive representation for local image descriptors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 506–513, 2004.

22. Iasonas Kokkinos, Michael Bronstein, and Alan Yuille. Dense Scale Invariant Descriptors for Images and Surfaces. Research Report RR-7914, INRIA, March 2012.

23. A. Kovnatsky, M. Bronstein, A. Bronstein, and R. Kimmel. Photometric heat kernel signatures. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 616–627, 2011.

24. M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *International Conference on Computer Vision*, pages 1482–1489, 2005.

25. Bruno Lévy. Laplace-Beltrami Eigenfunctions: Towards an Algorithm that Understands Geometry. In *IEEE International Conference on Shape Modeling and Applications*, page 13, June 2006.

26. H. Ling and D. Jacobs. Deformation invariant image matching. In *International Conference on Computer Vision*, pages 1466–1473, 2005.

27. H. Ling and D. Jacobs. Shape classification using the inner-distance. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 29(2):286–299, 2007.

28. H. Ling, X. Yang, and L. Latecki. Balancing deformability and discriminability for shape matching. In *European Conference on Computer Vision*, 2010.

29. R. Litman and AM. Bronstein. Learning spectral descriptors for deformable shape correspondence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(1):171–180, Jan 2014.

30. D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

31. K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 10(27):1615–1630, 2005.

32. J. Morel and G. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

33. F. Moreno-Noguer. Deformation and illumination invariant feature point descriptor. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1593–1600, 2011.

34. F. Moreno-Noguer and P. Fua. Stochastic exploration of ambiguities for nonrigid shape recovery. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(2):463–475, Feb 2013.

35. J. Philbin, M. Isard, J. Sivic, and A. Zisserman. Descriptor learning for efficient retrieval. In *European Conference on Computer Vision*, pages 677–691, 2010.

36. U. Pinkall and K. Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics*, 2(1):15–36, 1993.

37. Dan Raviv, Michael M. Bronstein, Nir Sochen, Alexander M. Bronstein, and Ron Kimmel. Affine-invariant diffusion geometry for the analysis of deformable 3d shapes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

38. M. Reuter, F. Wolter, and N. Peinecke. Laplace-beltrami spectra as 'shape-dna' of surfaces and solids. *Computer Aided Design*, 38(4):342–366, 2006.

39. R. Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Eurographics Symposium on Geometry Processing*, pages 225–233, 2007.

40. J. Sanchez, J. Ostlund, P. Fua, and F. Moreno-Noguer. Simultaneous pose, correspondence and non-rigid shape. In

IEEE Conference on Computer Vision and Pattern Recognition, pages 1189–1196, 2010.

41. E. Serradell, P. Glowacki, J. Kybic, F. Moreno-Noguer, and P. Fua. Robust non-rigid registration of 2d and 3d graphs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

42. L. Shi, Y. Yu, and N. Belland W.-W. Feng. A fast multigrid algorithm for mesh deformation. *ACM SIGGRAPH*, 25(3):1108–1117, 2006.

43. N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Transactions on Image Processing*, 7(3):310–318, 1998.

44. J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Eurographics Symposium on Geometry Processing*, pages 1383–1392, 2009.

45. F. Tang, S. Hwan Lim, N. Chang, and H. Tao. A novel feature descriptor invariant to complex brightness changes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2631–2638, 2009.

46. E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 32(5):815–830, 2010.

47. L. Torresani, V. Kolmogorov, and C. Rother. Feature correspondence via graph matching: Models and global optimization. In *European Conference on Computer Vision*, pages 596–609, 2008.

48. Eduard Trulls, Iasonas Kokkinos, Alberto Sanfeliu, and Francesc Moreno-Noguer. Dense segmentation-aware descriptors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

49. Eduard Trulls, Stavros Tsogkas, Iasonas Kokkinos, Alberto Sanfeliu, and Francesc Moreno-Noguer. Segmentation-aware deformable part models. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

50. A. Vaxman, M. Ben-Chen, and C. Gotsman. A multiresolution approach to heat kernels on discrete surfaces. *ACM SIGGRAPH*, 29(4):121, 2010.

51. A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008.

52. A. Vedaldi and S. Soatto. Features for recognition: Viewpoint invariance for non-planar scenes. In *International Conference on Computer Vision*, pages 1474–1481, 2005.

53. Z. Wang, B. Fan, and F. Wu. Local intensity order pattern for feature description. In *International Conference on Computer Vision*, pages 603 –610, 2011.

54. P. Wesseling. *An Introduction to Multigrid Methods*. John Wiley & Sons, 2004.

55. S. Yanowitz and A. Bruckstein. A new method for image segmentation. *Computer Vision, Graphics, and Image Processing*, 46(1):82–95, 1989.

56. A. Yezzi. Modified curvature motion for image smoothing and enhancement. *IEEE Transactions on Image Processing*, 7(3):345–352, 1998.