RSILC: Rotation- and Scale-Invariant, Line-based Color-aware descriptor☆

Sema Candemir *, Eugene Borovikov, K.C. Santosh, Sameer Antani, George Thoma

Lister Hill National Center for Biomedical Communications U. S. National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

ARTICLE INFO

ABSTRACT

Modern appearance-based object recognition systems typically involve feature/descriptor extraction and matching stages. The extracted descriptors are expected to be robust to illumination changes and to reasonable (rigid or affine) image/object transformations. Some descriptors work well for general object matching, but gray-scale key-point-based methods may be suboptimal for matching line-rich color scenes/objects such as cars, buildings, and faces. We present a Rotation- and Scale-Invariant, Line-based Color-aware descriptor (RSILC), which allows matching of objects/scenes in terms of their key-lines, line-region properties, and line spatial arrangements. An important special application is face matching, since face characteristics are best captured by lines/curves. We tested RSILC performance on publicly available datasets and compared it with other descriptors. Our experiments show that RSILC is more accurate in line-rich object description than other well-known generic object descriptors.

1. Introduction

Feature matching is an essential component of many modern computer vision applications, including near-duplicate detection [1], stereo correspondence [2], 3D modeling [3], image stitching [4], as well as face alignment and matching [5–7]. Scene and object matching methods in digital images can be roughly divided in the following major groups by the density of the image features they extract and use:

dense descriptor methods [8–10] tend to use all image information and often assume that all pixels in the image are equally important. Hence, they may be computationally expensive and require a high degree of correlation between the probe and gallery images. Typically, such methods are not very accurate given large variations in object pose, scale, and illumination.

sparse descriptor methods use non-dense image features (e.g., edges [11]) and/or various key-spots [12–14]. They are relatively robust to variations in pose, size, orientation, and lighting of the query image with respect to the objects in the gallery. They provide a sparse representation for objects and high-speed matching on the key locations that need to be automatically determined, which calls for some sort of a saliency map [15].

This implicit methodology division prompted some well-performing *hybrid* techniques that include features from both categories and typically fuse them in a weighted features ensemble [16–18], optimized for a particular application [19].

In content-based image retrieval, object detection, and face recognition (FR) in an unconstrained environment (typically outside of the studio), sparse descriptors are generally preferred because they are often more robust to deformations and lighting variations than dense feature methods. Key-spot-based matching involves detecting the key-spots, building local descriptors for each key-spot, and finding an aggregate distance of best matches. For pose- and lighting-robust matching, the descriptors should be robust to geometrical variations such as translation, rotation, scaling (and if possible to affine/projective transformations and photometric variations such as illumination direction, intensity, colors, and highlights. The selection of a robust key-spot (e.g., a point, a line, or a corner) depends on the image collection and the application.

In generic object matching, the coarse features from key-points may be adequate to find suitable matches. However, in line-rich scenes, some dominant lines on objects may provide more stable and discriminative features than key-points. Such line-rich scenes and objects are omnipresent and can be natural (e.g., landscapes, plants, animals, humans) or artificial (e.g., cities, cars, house exteriors and interiors, office spaces). Stable (but not necessarily rigid) characteristic lines in them can be used as good key-spot candidates, promising a more stable object matching ability than key-points can.

Human face matching/recognition (FM/FR) is an important special case of object matching that has been an active research and development area in academia and industry because of the wide variety of real-world applications, such as surveillance, visual authentication, human–machine interface, criminal identification, and commercial applications. It identifies individuals from face images or video sequences

using computer vision and machine learning algorithms. The general procedure for the appearance-based face image retrieval (FIR) systems consists of detecting the faces, extracting the facial information, and comparing a query face descriptor with those in a database [7,20–24].

To deal with the variation in face appearance (e.g., unknown head poses, unexpected facial expressions, and unpredictable lighting), the extracted features are expected to be robust to illumination changes, distortion, and scaling. One can certainly use key-point-based descriptors (e.g., SIFT [12], SURF [13], ORB [14], but because faces are line-rich objects, it may be beneficial to introduce the notion of key-lines and their descriptors for better matching.

According to psychophysics and neuroscience studies, line-rich features, such as face outline, eyes, mouth, and hair, are most important for perceiving and remembering faces [25] by humans. Another study has investigated the importance of facial features for automatic face recognition [26] by extracting facial landmarks. Experimental results indicate that these facial features are indeed important for face identification. Several other studies [6,11,27] showed that the most informative face characteristics appear to come from the face lines that can model face features in a very intuitive, human-perceptible form.

Consider the face images shown in Fig. 1.a–b. The prominent characteristic parts are marked by lines/curves, whose local regions and their spatial arrangements on a face can be used for robust matching. The human-perceptible important face lines overlap very well with the machine-computed edge maps of the faces on the CalTech set [28], whose cumulative distribution is also shown. The lines are mostly located on the prominent face characteristics (landmarks such as eye, mouth, nose, and face shape), which are the discriminative locations of a face (Fig. 1.c). All these studies and illustrations indicate that lines with their descriptors can provide more stable recognition features for face matching.

We propose a general-purpose key-line descriptor that is color aware, invariant to rotation/scale, and is robust to illumination changes. To increase the discriminative power of the descriptor, we combined color/texture information of the local regions and added the relational information of the other key-lines. We tested our descriptor matching power on publicly available datasets containing unconstrained images of faces and general objects. We compared the new descriptor to well-known descriptors (in the same test-bed system), and our experimental results show that the RSILC descriptor is robust to rotation, scale, reflection and illumination and produces more accurate matches in face and object retrieval applications.

2. Relevant publications

Many different techniques for modeling local image regions have been developed. Scale-Invariant Feature Transform (SIFT) [12,29] is one of the most robust key-point descriptors among the local feature descriptors with respect to different geometrical changes [30]. It detects notable and stable key-points for images at different resolutions and

produces scale- and rotation-invariant descriptors for robust matching. Several papers have been published on SIFT-based face recognition [31–33]. Although SIFT originally was designed for gray-scale images, there are several extensions to make use of the color information in the descriptors [34–36]. One of the successful attempts is colored SIFT (CSIFT) [36] which embeds the color information through the gradient of color invariants instead of using gray-scale gradients as in conventional SIFT. Defining the descriptor in color space makes the descriptor more robust with respect to color variations.

Another well-known key-point descriptor is Speeded Up Robust Features (SURF), which provides a quicker way to detect key-points and compute descriptors that are rotation and scale invariant as well as robust to illumination changes [37]. SURF is less accurate than SIFT, but it has been successfully used in many practical applications, including face/components matching [38,39].

The mentioned key-point descriptors (e.g. SIFT and SURF), being robust to various affine transformations and lighting, are widely used for object detection and recognition. However, they typically contain information that is local to their key-points, which prompts some false-positive correspondences when performing many-to-many matches. This problem could be remedied by considering key-point spatial relationships, (e.g., having each descriptor record other key-points' azimuth angles much like shape context [40]), hence capturing not only local context of each key-spot but also their global spatial relationships. Knowing the usefulness of spatial relations in image understanding [41,42], several state-of-the-art methods have been reported together with the use of application-dependent local features. For example, the authors in [43,44] integrate spatial distribution of key-points by using shape context with texture features for food classification. In a similar fashion, spatial relations between the visual primitives (such as circles and corners) are integrated with statistical shape features for graphics recognition [45].

The Pyramid of Histograms of Orientation Gradients (PHOG) descriptor represents an image by its local shape and the spatial layout of shape information [46]. The local shape is modeled by a histogram of edge orientations. The spatial layout is represented by tiling the image into the regions at multiple resolutions. The final descriptor vector is the concatenation of histograms at each resolution. The descriptor is robust to scaling as long as the object position and orientation remain the same, but it is rotation dependent and color-blind.

As an alternative to key-points, another important set of features for object matching can be collected from edges, which provide the advantages of a lesser demand on the storage space and a lower sensitivity to illumination changes. Gao and Leung [11] describe a face recognition method using line edge maps (LEMs). The system extracts the lines from the edge map of face images and compares their similarity using the Hausdorff distance. LEM produces fast and reliable matching on aligned faces but does not use the region around the lines, which contains intensity information that helps to discriminate the lines and reduce mismatches. Gao and Qi [47] extend the LEM approach by considering corner points along the edge lines and show their method's robustness to scale as well as the superior one-image-per-person retrieval capability compared to the eigenfaces [48]. Deboeverie et al. [6] combined the curve edge map with the relative positions and intensity information around the curves. This system uses the orientation of the main axis of the curve segments for the first match. Then, it considers the histograms of inner and outer sides of the curve and relative positions of curve segments to refine the matching stage. However, this method lacks color information for the local regions.

Liu et al. [16] propose SIFT flow to align an image to its nearest neighbors in a photo gallery. This hybrid method matches densely sampled, pixel-wise SIFT features between two images while preserving (sparse) spatial discontinuities, matching a query object located at different parts of the scene. Experiments show that the proposed approach robustly aligns complex scene pairs containing significant spatial differences. The applications include single image motion field prediction/
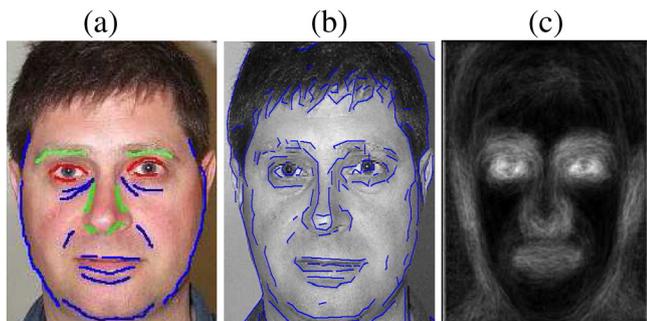
Fig. 1. Line/curve features that may be characteristic of a specific face: (a) human-perceptible, (b) line edge map (LEM) [11], (c) average line map of faces in the CalTech set [28].

synthesis via object transfer, satellite image registration, and face recognition.

Motivated by the discriminative ability of (sparse) shape information and (dense) local patterns in object recognition, Nguyen et al. [17] propose a window-based (hybrid) object descriptor that integrates both cues: contour templates representing object shape are used to derive a set of key-points at which local appearance features are extracted. The key-points are located via template matching that uses both spatial and orientation information. An object descriptor is formed by concatenating the non-redundant local binary pattern (NR-LBP) features from all key-points to encode the shape as well as the appearance of the object. The experimental results suggest that the proposed descriptor can effectively describe non-rigid articulated objects and improve the detection rate compared to other state-of-the-art object descriptors.

Satpathy et al. [18] propose two sets of edge-texture features for object recognition: Discriminative Robust Local Binary Pattern (DRLBP) and Ternary Pattern (DRLTP). These hybrid features fuse edge (sparse) and texture (dense) information in a single representation, and they appear to be robust to image variations caused by intensity inversion and discriminative to the image structures within the histogram block, as demonstrated by experiments on seven datasets.

Some recent papers present an interesting approach to line-rich scene matching [49,50]. This approach works with gray-scale images and uses rotation- and scale-invariant key-lines and log-polar descriptors. The authors described their promising methodology, giving a few examples with line-rich scenes but showing no large-scale experiments. The approach is quite similar to ours. Therefore, we highlight the major differences, as RSILC: 1) uses *color* information in the chroma bands of YC_bC_r color space, 2) uses local gradients histogram from the *equalized* intensity band, and 3) uses spatial information across all key-lines to make descriptor matching more accurate.

3. RSILC descriptor

Computing our key-line descriptor involves: 1) extracting key-lines by applying *oriented line filters* to the image edge maps, and 2) sampling the line direction oriented circular histograms around each key-line, including both local (texture, color) and global (inter-line spatial) information.

3.1. Key-lines

To extract key-lines from the studied image, we use convolution of *line filters* with an edge image, as reported in [49,50]. Formally, the whole process is composed of constructing line filters in different orientations and their convolution with an edge image. A set of line filters in a normal Gaussian 1D distribution are defined, which are oriented in different directions in the range $[0, \pi)$. In generic form, we can express such a set as $\mathcal{F} = \{F_\theta\}$, $\theta_k = \frac{k\pi}{6}$ and $k = 0, 1,...,5$, where θ_k represents an orientation value associated with a specific filter F. In Fig. 2, line filters that are oriented in six different directions are shown. These line filters F_θ are then convolved with an edge image I_E obtained via Canny edge detector [51]. The convolution results in a set
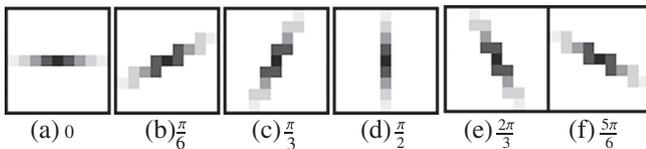
of key-lines from each line filter. Considering a set $\mathcal{F}$ of line filters, the complete set $\mathcal{L}$ of extracted key-lines can be expressed as

$$\mathcal{L} = F_\theta * I_E \quad \forall \theta \in [0, \pi) = \{<line_{\theta,l}>\}_{l=1,...,L}, \qquad (1)$$

where each extracted key-line carries its orientation value θ (that is associated with its line filter) and the total number of key-lines L that varies from one image to another. Note that small lines can be considered as noise and omitted. Fig. 3 shows key-lines that are extracted from a face image.

3.2. Key-line features

In this section, we describe how key-line features are computed to include both local and relational features. Local features are computed independently while relational features take other key-lines into account. We model each key-line with a circle of radius r. Fig. 4 shows the circles corresponding to key-lines from face images with different poses. Because the size of the radius uses the length of the key-line, circle sizes vary. Given a complete circle, we sub-divide it into equal sectors $s_i \in \mathcal{S}$. For each sector, local and relational features are computed and concatenated to build a key-line descriptor.

3.2.1. Local features

Each key-line is represented by gradient and color features that characterize the local texture and intensity distribution in every sector of the associated circle. For intensity information, YC_bC_r color space is used, which is robust to illumination changes and perceptually uniform compared to other color spaces [52]. The normalized Y-band is used to compute the gradient direction and gradient magnitude at each pixel. The color information is calculated from each color band (CB) separately. For each sector s_i, feature histograms are computed as follows,

$$
\begin{aligned}
h_{s_i}^{|\nabla I|} &= hist\left(|\nabla I|_{s_i}^p\right), \; \forall \; p \in s_i, \\
h_{s_i}^{\psi(\nabla I)} &= hist\left(\psi(\nabla I)_{s_i}^p\right), \; \forall \; p \in s_i, \\
h_{s_i}^{CB} &= hist(I(p)), \; \forall \; p \in s_i,
\end{aligned}
\qquad (2)
$$

where $hist(\star)$ is the histogram function and populates the values $\star$ into bins; $|\nabla I|_{s_i}^p = \sqrt{\left(\frac{\partial I(p)}{\partial x}\right)^2 + \left(\frac{\partial I(p)}{\partial y}\right)^2}$ is the gradient magnitude; $\psi(\nabla I)_{s_i}^p = atan(\frac{\partial I(p)}{\partial y} / \frac{\partial I(p)}{\partial x})$ is the gradient direction; $\frac{\partial I(p)}{\partial \star}$ is the gradient component in $\star$ direction; and $p = \{x, y\}$ is the image pixel. To maintain the rotation invariance of $h_{s_i}^{\psi(\nabla I)}$, the gradient angles are normalized with respect to the key-line orientation before the histogram computation. With the
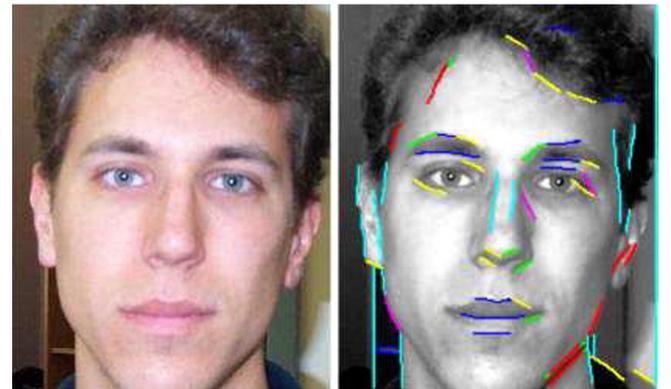
Fig. 3. An example face image from the CalTech dataset [28] and its key-lines. The key-lines are extracted from the convolution of line filters at different orientations illustrated in Fig. 1. The color indicates the orientation of key-lines: blue, 0; green, $\frac{\pi}{6}$; red, $\frac{\pi}{3}$; cyan, $\frac{\pi}{2}$; magenta, $\frac{2\pi}{3}$; yellow, $\frac{5\pi}{6}$.

(a) 0   (b) $\frac{\pi}{6}$ (c) $\frac{\pi}{3}$ (d) $\frac{\pi}{2}$ (e) $\frac{2\pi}{3}$ (f) $\frac{5\pi}{6}$

Fig. 2. Line filters that are defined as an angle oriented 1D Gaussian distribution. These filters are derived in the range $[0, \pi)$ with the step of $\frac{\pi}{6}$.
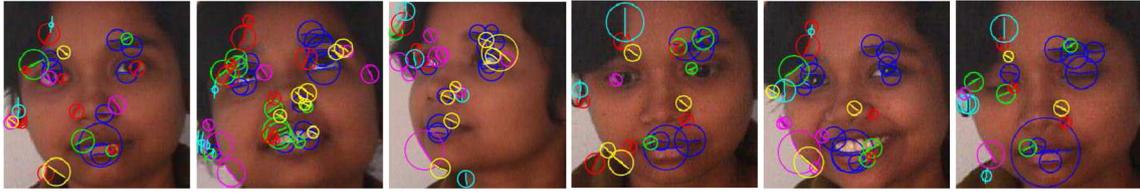
Fig. 4. Examples illustrating the detected key-lines along with their circles on a face of same person in several different poses. The color indicates the orientation of each key-line.

three separate histograms for color bands viz. Y, C_b, and C_r, the complete feature histogram in any particular sector can be expressed by concatenating them, $H_{loc_{s_i}} = [h_{s_i}^{|\nabla I|} \| h_{s_i}^{\psi(\nabla I)} \| h_{s_i}^{Y} \| h_{s_i}^{Cb} \| h_{s_i}^{Cr}]$. Considering all sectors in the specified circle for any key-line l, the complete local feature histogram can be expressed by concatenating them,

$$F_{loc}^{l} = \left[H_{loc_{s_1}} \| H_{loc_{s_2}} \| \dots \| H_{loc_{s_N}} \right]. \tag{3}$$

Such a concatenation of local feature histograms is more discriminant than the feature histogram computed from the whole circle at once [49,50].

To satisfy the rotation-invariant property, we follow counterclockwise histogram concatenation, which starts from the reference key-line orientation defined in Eq. (1). Fig. 5 shows a graphical illustration of rotation-invariance property of the descriptor. In this example, we have eight sectors where concatenation starts from the orientation angle of that particular reference key-line. Starting from the orientation angle of the key-line provides dynamic change in the sector indexing. As a consequence, the feature vector after concatenation remains the same for all key-lines in Fig. 5.a. In Section 3.3, real-world examples are considered to confirm the rotation-invariant property.

3.2.2. Relational features

Alone, local features for each key-line do not offer information about spatial organization of the remaining key-lines. This may cause some false matches. As an example, a key-line associated with the right eye of the query image could be matched with the key-line associated with the left eye of the database image. Therefore, relational

information is incorporated to the local feature histograms in each sector. Consider a sector s of a reference key-line l. We define two features to compute the relationships between the key-lines: orientation angle differences between l and the other key-lines $l' \in s_i$, and the presence of key-lines in that sector s_i with respect to l.

The orientation angle difference $\Delta\theta$ between l and l' can be computed as $\Delta\theta = |\theta - \theta'| \ \forall \ l' \in s$. Based on this, a histogram can be computed as

$$h_{s_i}^{\Delta\theta} = hist(|\theta - \theta'|). \tag{4}$$

where $hist(\star)$ is the histogram function and populates the values $\star$ into bins; the number of bins is defined as $\frac{k\pi}{6}$ and $k = [0, \dots, 5]$.

Additionally, the presence of key-lines is computed by taking the proportion of key-lines l' that are visible from l in s_i. To make the descriptor more discriminant, we further sub-divide s_i into annular sectors $a \in s_i$ and compute the histogram,

$$h_{s_i}^{\#l'} = hist(\#l')) \forall \ l' \in s_i, \tag{5}$$

where $hist(\star)$ is the histogram function and populates the values $\star$ into bins; the number of bins is $\{a_j\}_{j=1,\dots,4}$ and the size of the annular sector depends on the radius r of the circle belonging to the reference key-line.

In both relational features, histograms are normalized by using the total number of key-lines that appear in that sector. To graphically illustrate the relational feature histograms, we refer to Fig. 6. In one particular sector s_i, both histograms are concatenated to build the relational histogram $H_{rel_{s_i}} = [h_{s_i}^{\Delta\theta} \| h_{s_i}^{\#l'}]$. Considering all sectors for a single key-

(a)

(b)
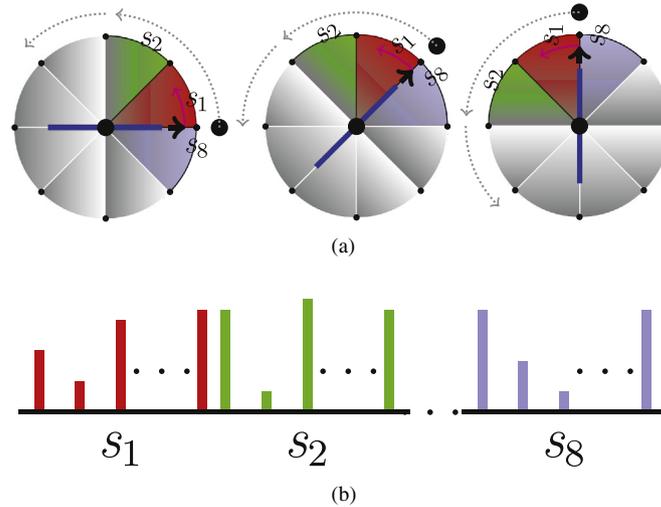
Fig. 5. An illustration of how local histograms are computed in each sector (where sector: $\{s_i\}_{i=1,\dots,N}$ and $N = 8$) using local features. (a) For an encircled key-line at three different angles: 0, $\frac{\pi}{4}$ and $\frac{\pi}{2}$ radians, we provide color in each sector, and (b) concatenation of all local histograms to produce the feature vector. While concatenating, we consider the orientation angle of key-line to preserve the rotation invariance.

(a) Key lines in sector s_1 (b) Relational histograms   (c) Concatenate both
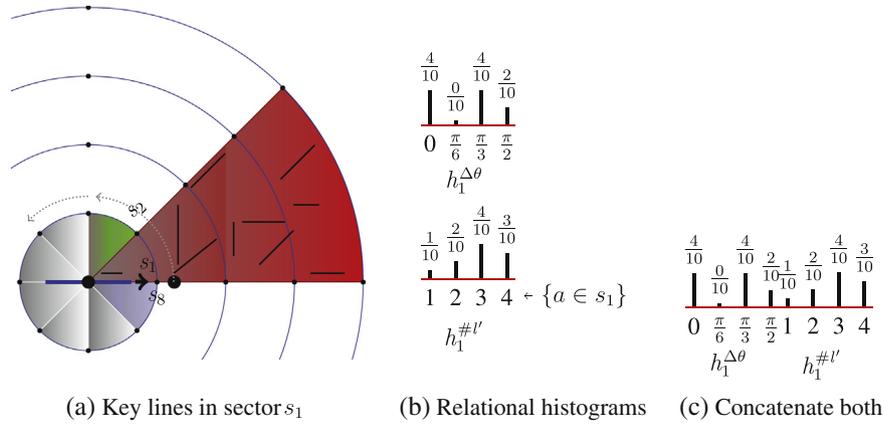
Fig. 6. An illustration of how relational histograms are computed in each sector. For an encircled key-line at 0 radians, we provide (a) the first sector where (b) relational histograms are computed. For simplicity, few key-lines are shown and four annular sectors are provided.

line l, the complete relational feature can be computed by concatenating all histograms,

$$F_{rel}^l = \left[H_{rel_{s_1}} \left\| H_{rel_{s_2}} \right\| \ldots \left\| H_{rel_{s_N}} \right] . \tag{6}$$

Like the local histograms, to satisfy the property of rotation invariance, concatenation follows exactly similar sector indexing.

3.3. Rotation and scale invariance

RSILC descriptor is designed to be rotation and scale invariant. To show these properties, we take example images as shown in Fig. 7. Example key-line locations are shown on straight, rotated, and re-scaled faces. For illustration purposes, we use a single key-line to build an RSILC descriptor (cf. Section 3.2). Histograms (both local and relational) from all sectors inside the circle are concatenated in a counterclockwise fashion, where the starting point is indexed by the reference key-line orientation angle. To maintain the scale invariance property, the reference key-line is modeled with a circle (cf. Fig. 5). The radius of circle is computed by using key-line length that varies with image size. Fig. 8 shows the corresponding 1D feature vector of example images. The overlapping of feature vectors (histogram distributions) from the mouth regions of up-right and rotated faces as well as from the eye regions of up-right and scaled faces show that RSILC is rotation and scale invariant.

4. Descriptor matching

Following our feature description in Sections 3.2.1 and 3.2.2, a line-rich image I can be described with a descriptor of line features

$$dsc = \left\{ F^l \right\}_{l=1,\ldots,L} \tag{7}$$

where $F^l = [F_{loc}^l, F_{rel}^l]$, as in Eqs. (3) and (6), and L is the number of key-lines in the descriptor. To match descriptors from two different images dsc_1 and dsc_2, a symmetric distance function is defined as the average norm of left and right best match scores vectors (BMSV) m_{ij}

$$dist(dsc_1, \ dsc_2) = (||m_{12}|| + ||m_{21}||)/2 \tag{8}$$

$$m_{ij} = \left\{ \min_{l=1,\ldots,L} diff\left(F_i^k, \ F_j^l \right) \right\}_{k=1,\ldots,L_i} \tag{9}$$

$$diff\left(F_i^k, \ F_j^l \right) = \ ||F_i^k - F_j^l|| \tag{10}$$

where $i, \ j \in \{1, \ 2\}$, m_{ij} are the BMSV and $L_{i=1, \ 2}$ is the number of individual line features in each descriptor. Note that one has to compute two BMSV (left and right) to satisfy the distance symmetry requirement. In our experiments, we found that in some applications (e.g., matching faces), it may be sufficient to compute just one half of the match (e.g., left BMSV) without much loss in the accuracy. This makes our distance non-symmetric but cuts matching time by

(a)   (b)   (c)   (d)
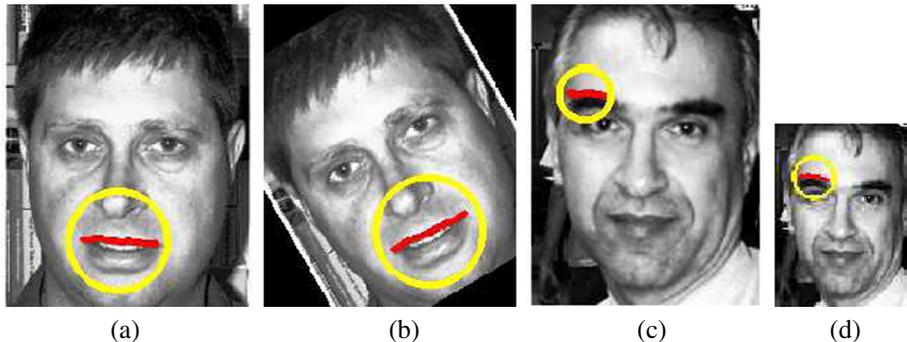
Fig. 7. a) An example straight face image from the CalTech set. b) A $\pi/6$ -rotated image of (a). c) Another face image from the CalTech set. d) A 1/2 re-sized image of (c).
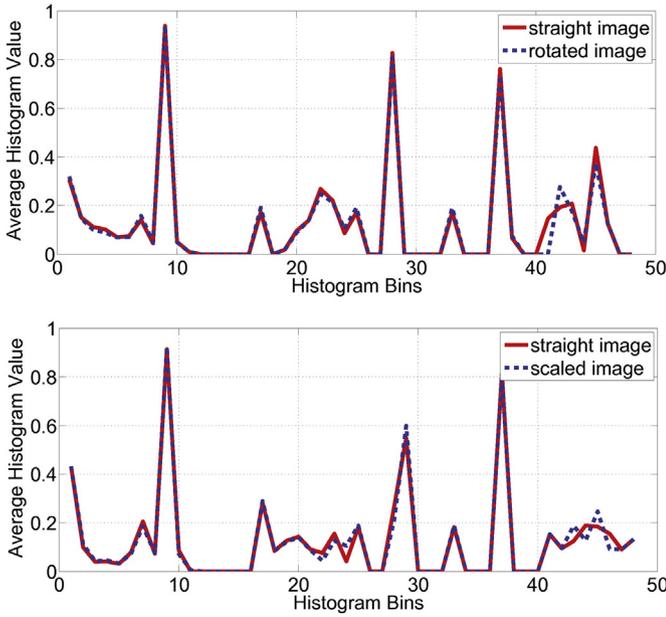
Fig. 8. a) Histogram distributions of region in Fig. 7.a–b. The red line belongs to the mouth region in the straight face image; the blue dotted line belongs to the mouth region in the rotated face image. b) Histogram distributions of regions in Fig. 7.c–d. The red line belongs to the eyebrow region of the face image in Fig. 7.c; the blue dotted line belongs to the eyebrow region of the 1/2-resized face image.

approximately half. These time savings can be used for ensuring RSILC invariance to image reflection (a feature that most key-point descriptors lack, e.g., SIFT and SURF) by redefining

$$diff\left(F_1^k, \ F_2^l\right) = \min\left(\left\|F_1^k - F_2^l\right\|, \left\|F_1^k - \hat{F}_2^l\right\|\right) \tag{11}$$

$$\hat{F} = \left[F_{i=\left(\frac{n}{2}+1\right),\dots,n}, \ F_{j=1,\dots,\frac{n}{2}}\right] \tag{12}$$

where $\hat{F}$ is a half-swapped feature vector (of even length n) computed to ensure the mirror flip matching invariance of the key-line descriptor; this effectively rotates the histogram sectors by the angle of π about the center of the key-line. One can skip the extra half-swap histogram match if the key-line orientation span is extended to the $[0, \ 2\pi]$ range (e.g., by computing the major intensity gradient direction at the line center), then *diff* can again be computed as in Eq. (10).

Although there certainly may be other descriptor matching methods, ours is consistent with most matching methods in the literature. Given a relatively small number of key-lines (compared to number of key-points) in a typical image and a good discriminative power of RSILC, we did not need to remove any outliers or perform descriptors cross-check.

5. Experiments

We design a set of experiments to evaluate the proposed descriptor. We first measure the retrieval performance of the descriptor on several face and generic datasets. We compare its performance with several other well-known descriptors on the same test-bed system. We also measure the performance change of descriptors to illumination variations and geometric transformations.

5.1. Test-bed system

We test the performance of the descriptors on a test-bed system that extracts the descriptor of the query image and compares it with the descriptors of database images. For this task, we follow a leave-one-out approach by selecting one face as a query image and comparing it

with the remaining images. According to the similarity measure, the top-k most similar images are listed as candidate matches. A hit-match is counted if the candidate object belongs to the same class of the query image. We define the hit rate for the top-k matches as

$$HitRate(k) = \frac{HitCount(k, Q)}{|Q|}, \tag{13}$$

where Q represents a set of query image, $HitCount(,)$ is a function that counts the successful top-k matches using the query set of size $|Q|$.

5.2. Datasets

Descriptor performances are measured using the following public datasets. Fig. 9 shows examples from these datasets.

The CalTech Faces set [28] consists of 450 frontal views of 29 people which are taken under varying lighting, expressions and complex background conditions.

The Indian Faces set [53] consists of frontal and left/right profile faces with varying facial expressions of 61 individuals under varying lighting conditions.

The ColorFERET set [54] consists of 2,413 facial images from 856 individuals with frontal and left/right profile faces, with/without glasses and with various facial expressions. In our experiment, a subset of the ColorFERET dataset is created with 500 images from 101 individuals.

The Labeled Faces in Wild (LFW) set [55] consists of celebrity faces collected from the Internet. The faces are captured in completely uncontrolled environments. Therefore, faces have large pose, illumination, expression, and facial hair variations. Some of them contain partial occlusions by glasses, mustaches, or hats.

The Amsterdam Library of Object Images (ALOI) set [56] is a generic object set that contains a large number of objects under different illumination directions and different viewpoints.

5.3. Parameter selection

Before measuring the retrieval performance of the descriptor, we investigate the effects of descriptor parameters to the matching performance. Our goal is to select parameters that yield optimal performances for both hit rate and execution time. We vary the parameters and conduct the matching experiments on the CalTech and Indian faces datasets.

5.3.1. Line filter size

The line filters are convolved with edge image to extract the key-lines (cf. Section 3.1). Increasing the filter size increases the computation time of the key-line extraction stage because of the convolution operation. To measure the performance of the descriptor with different line filter sizes, we set the size as k × k where $k = 9, 11, 13, 15$. We found that larger filters do not have a significant impact on hit rate. Therefore, in the experiments, we use 9 × 9 filters, which is also adequate to model all key-line orientations.

5.3.2. Radius of the circle

Each key-line is modeled with a circle of radius r (cf. Section 3.2). We set the radius as $r = k \times L$, where $k = 1/2, 2/3, 3/4, 1$, and L is the length of the key-line. We found that $r = 2L/3$ is adequate to characterize the local texture and intensity information around the key-line.

5.3.3. Number of sectors

To model the spatial information, we sub-divide the key-line circle into equal sectors and build the descriptor by concatenating the histograms of each sector. The number of sectors inside the circle is set to s,
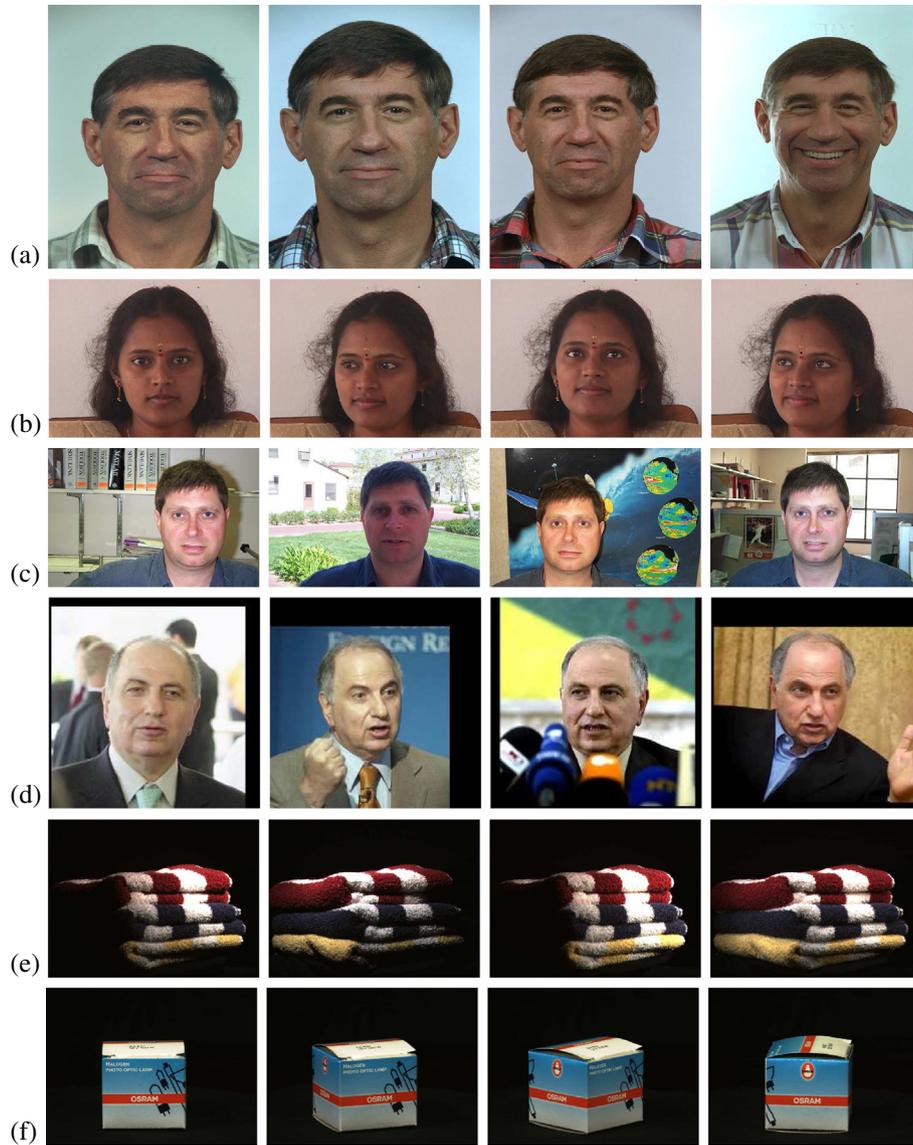
Fig. 9. Example images from the datasets used in our experiments. a) ColorFERET, b) Indian Faces, c) CalTech Faces, d) LWF set, e) ALOI illumination variance, f) ALOI view change.
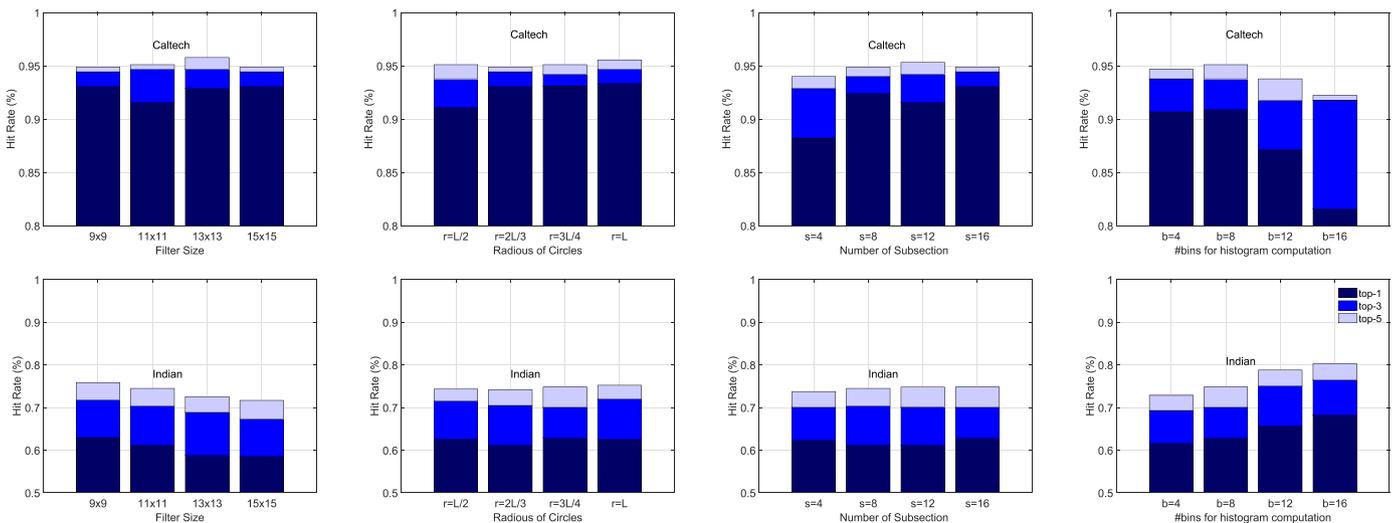
Fig. 10. The effects of the parameters of RSILC to the matching performance.

Table 1
Parameters used in the experiments.

Parameters for key-line descriptor	Value
Filter size	9×9 pixels
Radius of circle	2/3 of line length
No. of sector inside the circle	8
No. of bins for each sector (Eq. (2))	8

Parameters for Canny edge detection	Value
Gaussian filter σ	$\sqrt{2}$
Low hysteresis threshold	0.14
High hysteresis threshold	0.35

Average image size	Value
CalTech face dataset	$220 \times 185 \times 3$
Indian face dataset	$265 \times 265 \times 3$

where $s = 4, 8, 12, 16$. We found that $s = 8$ is adequate to model the spatial layout for the reported size of image.

5.3.4. Number of bins

In Eq. (2), the values inside the key-line circle are populated into bins. To observe the effect of the number of bins to descriptor discriminability, we set the bin size as $b = 4, 8, 12, 16$ and repeat the matching test. We found that $b = 8$ is best for the CalTech set, because retrieval performance decreases with a higher number of bins. However, for the Indian set (slightly larger images), we observed a higher hit rate with $b = 16$.

The graphs in Fig. 10 show the hit rates with respect to different parameter values. The parameter values used in the experiments are summarized in Table 1. Some of these parameters (e.g., Gaussian filter σ for Canny edge detection) vary with image size. Therefore, we also reported the average size of the test images.

5.4. Repeatability

One of the important stability measurements for a descriptor is its repeatability in terms of detected key-spot locations. To measure the repeatability, we follow similar test manner as in [29]. Each test image is (i) rotated in the range $[\pi/6]$ in $\pi/6$ intervals; (ii) re-scaled 1/2, 3/2, and 2 times the original size; and (iii) flipped in mirror-side or upside-down. Then, we extract the key-lines on the test image and on the transformed image. We expect that the key-lines are located at similar positions in the transformed image. Therefore, we compute where each key-line in the test image should appear in the transformed image by applying the same transformation to the key-lines in the test image. If the transformed key-line on the test image is L_T and the

extracted key-line on the transformed image is L_E, then the location accuracy of key-lines is computed with the Euclidean distance of the center coordinates of each L_T and L_E with the following equation:

$$d = \|C_{L_E} - C_{L_T}\| \tag{14}$$

where C_{L_E} is the center coordinates of L_E and C_{L_T} is the center coordinates of L_T. $d < \in$ means that L_E and L_T are $\in$-close, and their regions overlap. The repeatability is the percentage of overlapped key-lines. We measure the repeatability on the CalTech set with several different transformations. The graphs in Fig. 11 show the repeatability of the key-line locations. Each point in the graph corresponds to repeatability score of one image. The color of the points indicates the applied transformation. The x-axis is the image number on the CalTech set in which there are 450 face images. The y-axis shows the repeatability score. The median repeatability levels for the CalTech set are shown as color-coded solid lines. Experiment results show that RSILC is most invariant to mirror flip transformation, followed by rotation and scale transformation.

5.5. Baseline state-of-the-art descriptors

We compare the proposed descriptor with well-known baseline descriptors. We measure the performance of all descriptors with their best parameters on the same dataset and test-bed system (cf. Section 5.1). The baseline descriptors we compare ours to are as follows:

SIFT [12,29] is a well-known key-point descriptor that extracts the key-point in scale-space and models the region around the key-point with the statistics of gradient information. It is invariant to scale and rotation transformations and partially robust to illumination. The conventional SIFT descriptor does not use color information.
SURF [37], similar to SIFT, uses the distribution of gradient information for each key-point. The descriptor has smaller dimensions than SIFT, therefore, has relatively faster computation. It is invariant to scale and rotation and robust to lighting variations. It does not contain color information.
CSIFT [36] embeds the color information in conventional SIFT descriptor [29] through the gradient of color invariants instead of gray-scale gradients. Our experiments indicate that CSIFT is more successful than other descriptors at matching due to addition of color information.
PHOG [46] descriptor represents the image by its local and the spatial layout of the shape information. The local shape is represented by histogram of edge orientations; and spatial layout is represented by tiling the image into regions at multiple resolutions. The final descriptor vector is concatenation of histograms at each resolution. It is robust to scaling as long as the object position and orientation

(a) Repeatability for rotation transformations　　(b) Repeatability for scale transformations　　(c) Repeatability for flip transformations
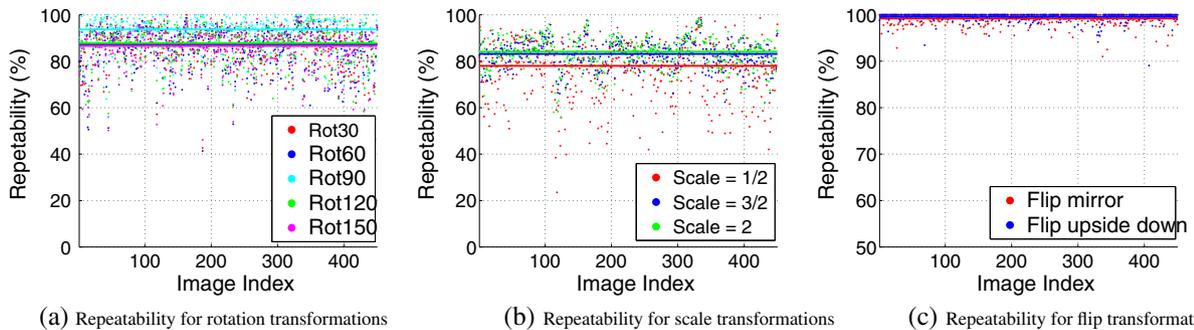
Fig. 11. The percentage of key-lines in each image that is repeatably detected at the same location in the transformed image. Each point corresponds to one image in the CalTech set. The color code shows the transformation cases. The average repeatability is indicated as line. The x-axis shows the image number, and the y-axis shows the repeatability score as a percentage of correctly repeated lines.
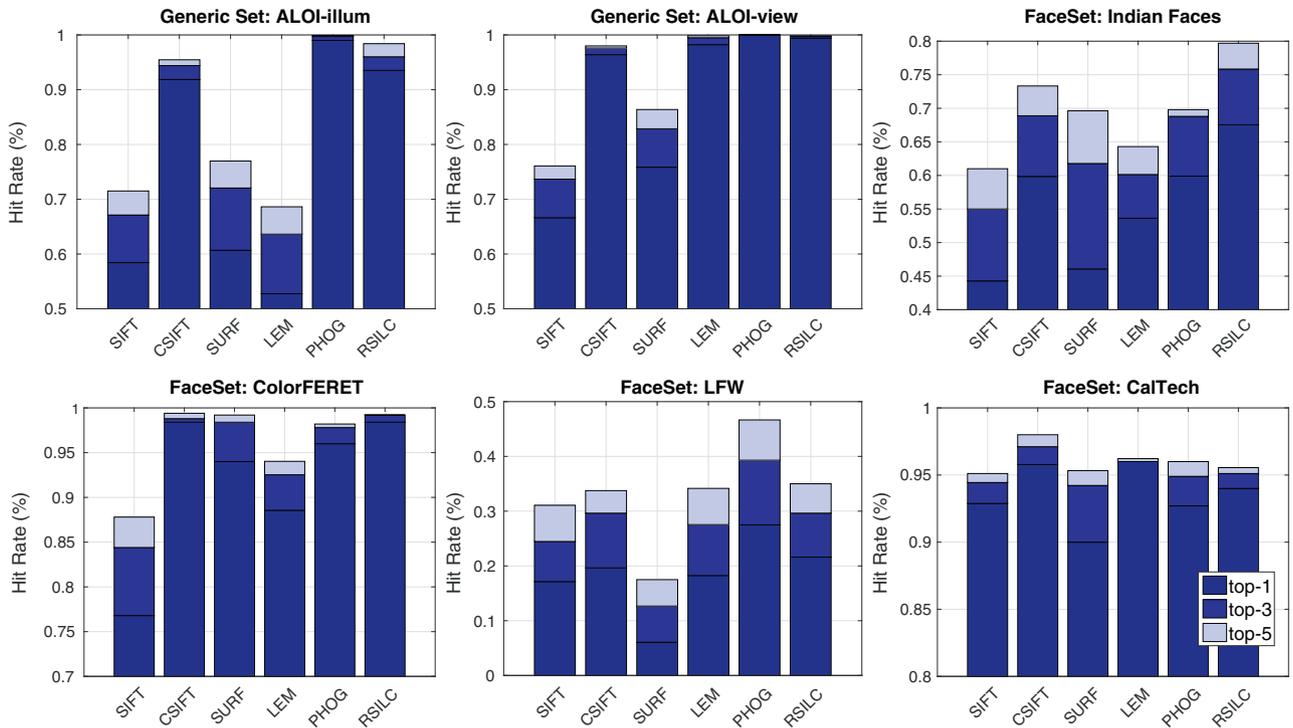
Fig. 12. Query performance of descriptors on generic and face datasets. Note that minimum scale values of the vertical axis are set different for illustration purpose.

remain the same. However, it is not rotation invariant and does not use color information.

LEM [11] is the only key-line descriptor among the comparison descriptors. Modeling with lines reduces the storage requirement and sensitivity to illumination change. However, ignoring the texture information of objects could decrease the matching performance. In the experiments, we use face matching system with LEMs [11], which extract the lines from the edge map of faces and compare the lines using Hausdorff distance. This basic system performs as well as more advanced face matching techniques, but it is not designed for geometric transformation variances.

5.6. Results

5.6.1. Comparison study

The bar graphs in Fig. 12 summarize the hit rate (%) of each descriptor on the test datasets. We found that the retrieval performance of RSILC is either the first or the second among the other descriptors for all test datasets. CSIFT is one of the best descriptors except on the Indian dataset. This descriptor uses color information, therefore it has better results compared to color-blind descriptors. LEM performance is higher than the other descriptors on the CalTech set, in which all images are frontal faces. However, LEM is not designed to deal with geometric transformations. Therefore, its matching accuracy decreases
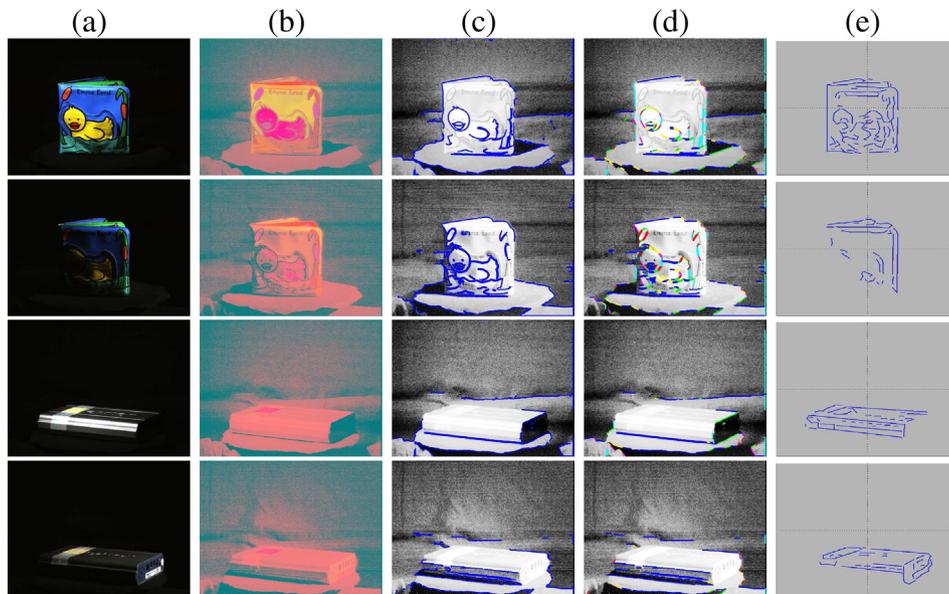
Fig. 13. Reaction to illumination change of the key-lines vs LEM. a) Objects in RGB space. b) Objects in YC_bC_r space. c) Edge maps computed in YC_bC_r space. d) Key-lines extracted from edge maps computed in YC_bC_r space. e) Edge maps computed from RGB space.

(a) Rotation $(\pi/6)$ Transformation   (b) Mirror-flip Transformation   (c) Scale $(\frac{1}{2} x)$ Transformation
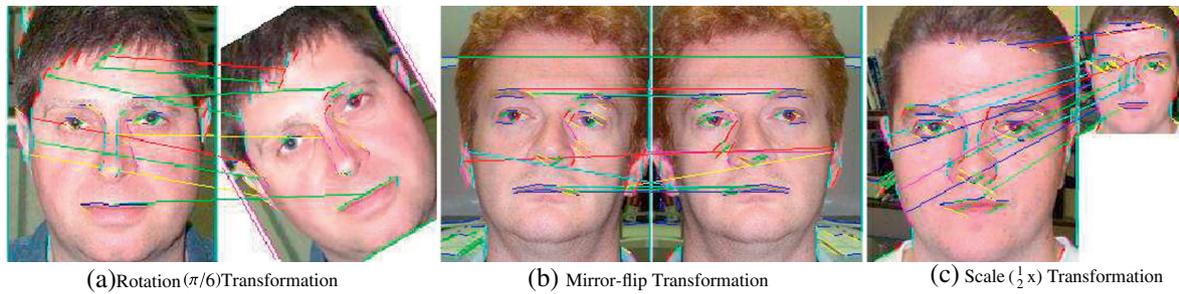
Fig. 14. An example showing the robustness of the RSILC descriptor to the (a) rotation $(\pi/6)$, (b) mirror-flip, and (c) scale $(\frac{1}{2}x)$ transformations. Extracted key-lines and corresponding matches are illustrated as color lines.

on a dataset with large pose variations (e.g., Indian Faces and ColorFERET) and performance is drastically lower if the query image is rotated and re-scaled (cf. Fig. 15 and Section 5.6.3). The PHOG descriptor produces comparable results on the face datasets and highest scores on the generic set but is color-blind and rotation variant. Therefore, in transformation experiments, this descriptor failed to perform accurate matching (cf. Section 5.6.3). Conventional SIFT and SURF have lower performance on both the face and generic sets compared to other descriptors.

In a controlled environment, all descriptors perform well, even with a basic test-bed system as in our case. However, the matching results on the LFW dataset show that a simple test-bed system is insufficient for face recognition in an unconstrained environment. It would be unfair to compare our scores with full face identification systems, which often have additional stages, such as the incorporation of age, gender, or facial expression information to improve matching accuracy. Because we propose a generic descriptor, we compare it with the other generic descriptors in the same test-bed system.

5.6.2. Robustness to illumination

All of the face datasets that we used in our experiments have illumination variances including other variances, such as view and facial expression. However, in the ALOI set, the illumination direction is systematically varied for each object in 24 configurations and all other parameters are held constant. Therefore, we found that this database is suitable to test the robustness of the descriptor to large illumination variance. As mentioned in Section 3.2.1, our descriptor uses YC_bC_r color space, which is more robust to illumination changes compared to other color spaces. Fig. 13 shows two example objects with different illumination directions. The RSILC descriptor computes the key-lines from the edge map of the objects and the edge boundary is computed from the Y-band of YC_bC_r color space. The visibility of objects is affected by high illumination variance in RGB space. On the other hand, the

illumination variance seems to be reduced in YC_bC_r space. Therefore, the edge maps extracted from the Y-band are less affected by illumination variance. The other line-based descriptor, LEM, extracts the edges from the RGB space. Although it is claimed that the LEM is robust to illumination changes because it only uses edge information, the edge maps in Fig. 13.e and the poor retrieval performance on the ALOI illumination set indicate that edges are affected by the high illumination variance. The retrieval accuracy of the conventional SIFT and SURF on this dataset are also lower (Fig. 12). It can be interpreted as, these descriptors are less robust to illumination variance compared to CSIFT, PHOG, and RSILC, which perform well in this set with a retrieval score (hit rate) higher than 90%.

5.6.3. Robustness to transformation

We investigate the retrieval performance of the descriptors to the following transformations: rotation, scaling, and mirror-flipping. We run experiments on the CalTech set. The query image is rotated by $\pi/6$, rescaled to 50% of its original size, mirror-flipped, and matched with straight images in the dataset. Fig. 14 shows examples from the experiment. We report the hit rates of descriptors as bar graphs in Fig. 15. According to the scoring, RSILC matching performance is not affected by rotation, scaling, or flip transformation, but SIFT, SURF, PHOG, and LEM performance suffered a significant drop when images were transformed. LEM and PHOG are not designed to cope with rotation transformation. Thus, their matching performance in the rotation case was less than 50%.

To quantify the degradation of retrieval performance with image transformation, we compare retrieval performance of RSILC on transformed images with the retrieval performance on non-transformed images. Table 2 lists the hit rates of the system on the CalTech set. There is a slight drop due to transformation at the top-1 hit rate. However, at top-3 and top-5, the hit rate is equal to that obtained with non-transformed query images.
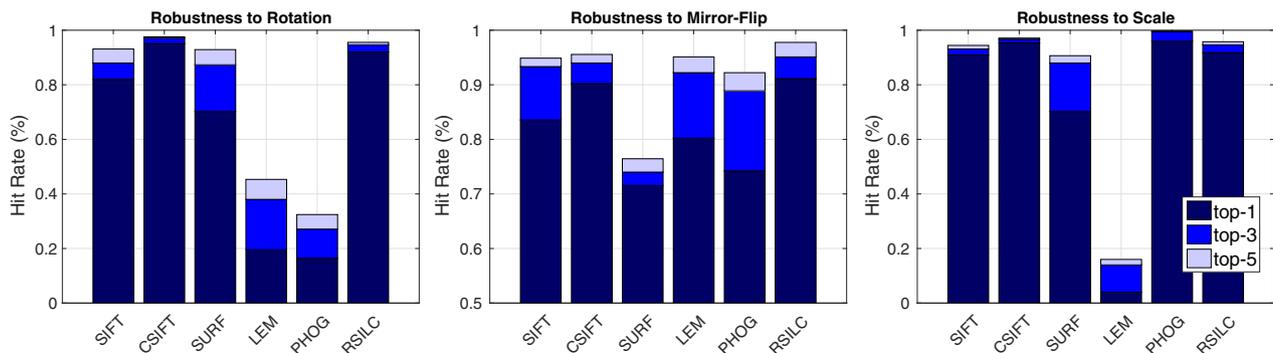
Fig. 15. Retrieval performance (hit rate) of descriptors with transformed query images and straight face images in the CalTech set.

Table 2
The effects of transformation on retrieval performance. The test is conducted on the CalTech set.

Query image	Hit rate		
	Top-1	Top-3	Top-5
Non-transformed	0.9400	0.9511	0.9555
π/6-rotated	0.9299	0.9467	0.9555
1/2-scaled	0.9178	0.9467	0.9578
Mirror-flipped	0.9111	0.9511	0.9777

5.7. Evaluation in a real-world system

We have successfully integrated RSILC in the face matching module (FaceMatch) [19] of the People Locator (PL) [57] service developed by NLM to help families reunite in disaster situations. PL photo collection is dynamic and large, containing hundreds of thousands of images. It typically has only one (seldom two) picture per person; therefore, FaceMatch cannot train any sophisticated person-specific model and is required to provide single-shot-based face matching capability while working with fairly low-quality images. Our experiments indicate that overall, RSILC provides a better top-1 hit rate face image retrieval accuracy ($HitRate$.RSILC = 0.96) than any other individual descriptor, including the rotation- and scale-invariant SIFT (0.94) and SURF (0.92), on the CalTech Faces, yielding only to the ensemble (an optimally weighted combination) of descriptors (SIFT, SURF, ORB, LBPH, HAAR) ($HitRate$.COMBO = 0.98), thus prompting a safe replacement of some of them with itself without any expected accuracy loss.

6. Conclusions

We introduced a new Rotation- and Scale-Invariant Line-based Color-aware descriptor (RSILC), which detects image key-lines and their circular regions, combining both local (intensity, color, gradient histograms) and global (line inter-positioning) information. It models the line-rich objects in a more intuitive and economical way than the popular key-point descriptors do. Moreover, RSILC has a better matching efficiency, since it produces more discriminative candidates to match in line-rich regions of interest. Aside from the mentioned features, RSILC is also mirror-flip invariant via symmetric key-line descriptor matching, something that other key-spot descriptors we tested cannot claim.

We performed a thorough investigation of the descriptor matching performance by conducting the following image and face retrieval experiments involving (i) generic line-rich objects and faces (an important special case of line-rich objects), (ii) controlled illumination variation using several direct lights, (iii) geometric transformation, including rotation, scale, and mirror-flip transformations. We compared our descriptor matching performance with well-known descriptors in the same test-bed system. The results indicated that overall RSILC is more accurate for line-rich object matching than the key-point descriptors. We successfully integrated RSILC in the real-world face image retrieval system FaceMatch (FM) and observed that RSILC is more accurate than any of the FM individual descriptors.

Our future work includes experiments with alternative edge detection methods (e.g., key-line filtering directly after the Sobel operator), various optimizations to RSILC computation and matching (e.g., reducing the individual descriptor size) utilizing multiple processing cores, and pushing batch matching to GPU. RSILC can be further expanded to piece-wise polynomial and curve-based region matching, which should expand the possible image matching application pool. Although RSILC produced good results in our out-of-plane rotation experiments, it was not designed to be invariant to affine or projective transforms, hence a natural extension would involve the study and introduction of such an invariance.

References

[1] W.-L. Zhao, C.-W. Ngo, Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection, IEEE Trans. Image Process. 18 (2) (2009) 412–423, http://dx.doi.org/10.1109/TIP.2008.2008900.
[2] A.S. Ogale, Y. Aloimonos, Robust contrast invariant stereo correspondence, IEEE International Conference on Robotics and Automation 2005, pp. 819–824, http://dx.doi.org/10.1109/ROBOT.2005.1570218.
[3] E. Borovikov, L. Davis, 3d shape estimation based on density driven model fitting, First International Symposium on 3D Data Processing Visualization and Transmission 2002, pp. 116–125, http://dx.doi.org/10.1109/TDPVT.2002.1024051.
[4] Z.-L. Yang, B.-L. Guo, Image mosaic based on sift, International Conference on Intelligent Information Hiding and Multimedia, Signal Processing 2008, pp. 1422–1425, http://dx.doi.org/10.1109/IIH-MSP.2008.335.
[5] M. Luo, D. Dementhon, X. Yu, D. Doermann, SoftCBIR: Object searching in videos combining keypoint matching and graduated assignment, 2006.
[6] F. Deboeverie, P. Veelaert, W. Philips, Face analysis using curve edge maps, Lecture Notes in Computer Science, Vol. 6979, Springer 2011, pp. 109–118.
[7] S.D. Lin, J.-H. Lin, C.-C. Chiang, Using gradient features from scale-invariant keypoints on face recognition, Int. J. Innov. Comput. Inf. Control 7 (4) (2011) 1639–1649.
[8] C.E. Jacobs, A. Finkelstein, D.H. Salesin, Fast multiresolution image querying, 22nd Annual Conference on Computer Graphics and Interactive Techniques, ACM, New York, NY, USA 1995, pp. 277–286.
[9] D.T. Nguyen, Z. Zong, P. Ogunbona, W. Li, Object detection using non-redundant local binary patterns, 17th IEEE International Conference on Image Processing 2010, pp. 4609–4612, http://dx.doi.org/10.1109/ICIP.2010.5651633.
[10] Z. Wang, J. Feng, S. Yan, H. Xi, Image classification via object-aware holistic superpixel selection, IEEE Trans. Image Process. 22 (11) (2013) 4341–4352.
[11] Y. Gao, M.K. Leung, Face recognition using line edge map, IEEE Trans. Pattern Anal. Mach. Intell. 24 (6) (2002) 764–779.
[12] D.G. Lowe, Object recognition from local scale-invariant features, IEEE International Conference on Computer Vision, Vol. 2 1999, pp. 1150–1157.
[13] H. Bay, T. Tuytelaars, L.V. Gool, SURF: speeded up robust features, European Conference on Computer Vision 2006, pp. 404–417.
[14] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: an efficient alternative to SIFT or SURF, IEEE International Conference on Computer Vision 2011, pp. 2564–2571.
[15] R. Cappelli, A. Franco, D. Maio, Gabor saliency map for face recognition, in: R. Cucchiara (Ed.), ICIAP, IEEE Computer Society 2007, pp. 443–447.
[16] C. Liu, J. Yuen, A. Torralba, J. Sivic, W.T. Freeman, SIFT flow: dense correspondence across different scenes, European Conference on Computer Vision, Springer 2008, pp. 28–42.
[17] D. Thanh Nguyen, P.O. Ogunbona, W. Li, A novel shape-based non-redundant local binary pattern descriptor for object detection, Pattern Recogn. 46 (5) (2013) 1485–1500, http://dx.doi.org/10.1016/j.patcog.2012.10.024.
[18] A. Satpathy, X. Jiang, H.-L. Eng, LBP-based edge-texture features for object recognition, IEEE Trans. Image Process. 23 (5) (2014) 1953–1964, http://dx.doi.org/10.1109/TIP.2014.2310123.
[19] E. Borovikov, S. Vajda, G. Lingappa, S. Antani, G. Thoma, Face matching for postdisaster family reunification, IEEE International Conference on Healthcare Informatics 2013, pp. 131–140, http://dx.doi.org/10.1109/ICHI.2013.23.
[20] X. Tan, S. Chen, Z.-H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, Pattern Recogn. 39 (9) (2006) 1725–1745.
[21] A.M. Martnez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. Pattern Anal. Mach. Intell. 24 (6) (2002) 748–763.
[22] B. Manjunath, R. Chellappa, C. von der Malsburg, A feature based approach to face recognition, Computer Society Conference on Computer Vision and Pattern Recognition, IEEE 1992, pp. 373–378.
[23] L. Wiskott, J.-M. Fellous, N. Kuiger, C. Von Der Malsburg, Face recognition by elastic bunch graph matching, Vol. 19IEEE, 1997. 775–779.
[24] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, Proceedings of the European Conference on Computer Vision, Springer 2004, pp. 469–481.
[25] G. Davies, H. Ellis, J. Shepherd, E.C. Carterette, M. Friedman, Perceiving and Remembering Faces (Cognition and Perception), Academic Press, 1981.
[26] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Trans. Pattern Anal. Mach. Intell. 15 (10) (1993) 1042–1052.
[27] Y. Liang, M. Song, L. Xie, J. Bu, C. Chen, Face sketch-to-photo synthesis from simple line drawing, Asia-Pacific Signal Information Processing Association Annual Summit and Conference 2012, pp. 1–5.
[28] CalTech frontal face dataset, http://www.vision.caltech.edu/archive.html/ (Online; accessed 10-October-2013).
[29] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[30] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.

[31] M. Bicego, A. Lagorio, E. Grosso, M. Tistarelli, On the use of sift features for face authentication, Conference on Computer Vision and Pattern Recognition Workshop, IEEE, 2006 (35-35).

[32] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, B.-L. Lu, Person-specific sift features for face recognition, International Conference on Acoustics, Speech and Signal Processing, Vol. 2, IEEE 2007, pp. 593–596.

[33] D.R. Kisku, A. Rattani, E. Grosso, M. Tistarelli, Face identification by sift-based complete graph topology, Workshop on Automatic Identification Advanced Technologies, IEEE 2007, pp. 63–68.

[34] A.A. Farag, A.E. Abdel-Hakim, Detection, categorization and recognition of road signs for autonomous navigation, Advanced Concepts for Intelligent Vision Systems 2004, pp. 125–130.

[35] K.E. van de Sande, T. Gevers, C.G. Snoek, Evaluating color descriptors for object and scene recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2010) 1582–1596, http://dx.doi.org/10.1109/TPAMI.2009.154.

[36] A. Abdel-Hakim, A. Farag, CSIFT: a sift descriptor with color invariant characteristics, Proceedings of the IEEE Computer Society Conference on Computer Vision and, Pattern Recognition, Vol. 2 2006, pp. 1978–1983, http://dx.doi.org/10.1109/CVPR. 2006.95.

[37] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (SURF), Comput. Vis. Image Underst. 110 (3) (2008) 346–359.

[38] D. Kim, R. Dahyot, Face components detection using SURF descriptors and SVMs, International Conference on Machine Vision and Image Processing, IEEE 2008, pp. 51–56.

[39] P. Dreuw, P. Steingrube, H. Hanselmann, H. Ney, Surf-face: face recognition under viewpoint consistency constraints, British Machine Vision Conference, 2009.

[40] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.

[41] J. Freeman, The modelling of spatial relations, Comput. Graph. Image Process. 4 (2) (1975) 156–171, http://dx.doi.org/10.1016/S0146-664X(75)80007-4.

[42] M. Bart, S. Ullman, Spatial context in recognition, Perception 25 (1996) 343–352.

[43] D.T. Nguyen, Z. Zong, P.O. Ogunbona, Y. Probst, W. Li, Food image classification using local appearance and global structural information, Neurocomputing 140 (2014) 242–251.

[44] Z. Zong, D.T. Nguyen, P. Ogunbona, W. Li, On the combination of local texture and global structure for food classification, IEEE International Symposium on Multimedia 2010, pp. 204–211, http://dx.doi.org/10.1109/ISM.2010.37.

[45] K.C. Santosh, B. Lamiroy, L. Wendling, Integrating vocabulary clustering with spatial relations for symbol recognition, Int. J. Doc. Anal. Recognit. 17 (1) (2014) 61–78, http://dx.doi.org/10.1007/s10032-013-0205-4.

[46] A. Bosch, A. Zisserman, X. Munoz, Representing shape with a spatial pyramid kernel, Proceedings of the 6th International Conference on Image and Video Retrieval, ACM 2007, pp. 401–408.

[47] Y. Gao, Y. Qi, Robust visual similarity retrieval in single model face databases, Pattern Recogn. 38 (7) (2005) 1009–1020.

[48] M. Turk, A. Pentland, Face recognition using eigenfaces, IEEE Computer Society Conference on Computer Vision and, Pattern Recognition 1991, pp. 586–591, http://dx. doi.org/10.1109/CVPR.1991.139758.

[49] Q.H.Z. Ye, W. YanJie, Rotation and scaling invariant feature lines for image matching, conference on mechatronic science, International Conference on Mechatronic Science, Electric Engineering and Computer 2011, pp. 1135–1138.

[50] Z. Ye, Q. Hongsong, W. Yanjie, Scene matching based on directional keylines and polar transform, IEEE International Conference on, Signal Processing 2010, pp. 857–860.

[51] J. Canny, A computational approach to edge detection, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 6 1986, pp. 679–698.

[52] R.-L. Hsu, M. Abdel-Mottaleb, A. Jain, Face detection in color images, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 696–706.

[53] The indian face database, http://vis-www.cs.umass.edu/vidit/IndianFaceDatabase/.

[54] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The feret evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.

[55] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Tech. rep., University of Massachusetts, Amherst, 2007.

[56] J.-M. Geusebroek, G.J. Burghouts, A.W. Smeulders, The Amsterdam library of object images, Int. J. Comput. Vis. 61 (1) (2005) 103–112.

[57] G. Thoma, S. Antani, M. Gill, G. Pearson, L. Neve, People locator: a system for family reunification, IT Prof. 14 (2012) 13–21.