A Study of the Region Covariance Descriptor: Impact of Feature Selection and Image Transformations

Hayden Faulkner*, Ergnoor Shehu*, Zygmunt L. Szpak*, Wojciech Chojnacki*
Jules R. Tapamo†, Anthony Dick* and Anton van den Hengel*
*School of Computer Science
The University of Adelaide, SA 5005, Australia
Email:{hayden.faulkner, ergnoor.shehu, zygmunt.szpak, wojciech.chojnacki,
anthony.dick, anton.vandenhengel}@adelaide.edu.au
†School of Engineering
University of KwaZulu-Natal, Durban 4041, South Africa
Email: tapamoj@ukzn.ac.za

***Abstract*—We analyse experimentally the region covariance descriptor which has proven useful in numerous computer vision applications. The properties of the descriptor—despite its widespread deployment—are not well understood or documented. In an attempt to uncover key attributes of the descriptor, we characterise the interdependence between the choice of features and distance measures through a series of meticulously designed and performed experiments. Our results paint a rather complex picture and underscore the necessity for more extensive empirical and theoretical work. In light of our findings, there is reason to believe that the region covariance descriptor will prove useful for methods that perform image super-resolution, deblurring, and denoising based on matching and retrieval of image patches from an image dictionary.**

I. Introduction

A modern computer vision pipeline for generic image classification and recognition consists of three broad conceptual steps. The first step involves selecting suitable image descriptors that capture essential characteristics of the class of images that need to be recognised. The second step involves the definition of a measure of similarity between feature descriptors. This is typically achieved by introducing a proper notion of distance between feature descriptors, with the intuitive interpretation that close feature descriptors are similar. The final step requires learning a classification rule that uses the feature descriptors and corresponding similarity measure to determine what the image represents. In this paper we discuss topics that touch upon the first two steps of the pipeline—the selection of suitable image descriptors and the definition of an appropriate distance. In this connection, we explore the *region covariance descriptor* which has proven to be very effective for a variety of computer vision tasks. Our aim is to gain a deeper understanding of this feature descriptor together with three accompanying measures of distance which are frequently utilised. We are motivated by the fact that despite the widespread use of the region covariance descriptor, its strengths and limitations are not well understood nor documented. Moreover, there appears to be no comprehensive appraisal of the impact that the choice of the measure of distance has on the utility of the descriptor. In fact, it seems that the choice of distance measure is problem and domain specific. On any given classification task, the distance measure is typically chosen on an ad-hoc basis, without any attempt to understand or characterise why a particular distance measure works better or worse. We take a first step toward addressing this knowledge gap by designing and conducting a series of targeted experiments which explore the strengths and limitations of the region covariance descriptor and three associated distance measures.

II. Related Work

A number of region descriptors have been proposed in the literature for a variety of tasks such as recognition and tracking. Among the simplest is the vector of pixel intensities [1]; however, raw image pixel intensities are poor descriptors because they are too variant to illumination and pose changes. Pixel intensities merely record the appearance of a scene. They do not model attributes of an image, and therefore cannot readily characterise what an image represents. Considerable research has focused on the development of useful region descriptors that take into account aspects such as colour, texture, and shape. A full account of the entire spectrum of descriptors is beyond the scope of this paper. Instead, we focus on the region covariance descriptor. Conceptually simple and with considerable expressive potential, the region covariance descriptor can capture aspects of colour, texture, and shape simultaneously by modelling the variations and correlations of various features in a region. In recent years it has been used in a variety of contexts, including tracking [2], detection and matching [3]–[9], as well as classification and recognition [10]–[14]. While some attempts have been made to characterise the performance of the region covariance descriptor [15], the evaluation has not been very comprehensive nor systematic. It has already been acknowledged that the choice of distance function matters [12], [16]. Nevertheless, limited attempts have been made to understand why a particular selection of features and a particular choice of distance metric works better for certain problems and not for others.

TABLE I: Description of Potential Features for $\phi(\mathbf{x})$

Notation		Description
xy	x	spatial x coordinate
	y	spatial y coordinate
rgb	r	red channel
	g	green channel
	b	blue channel
∂	$\lvert\mathbf{I_x}\rvert$	magnitude of first-order partial derivative in horizontal direction
	$\lvert\mathbf{I_y}\rvert$	magnitude of first-order partial derivative in vertical direction
∂^2	$\lvert\mathbf{I_{xx}}\rvert$	magnitude of second-order partial derivative in horizontal direction
	$\lvert\mathbf{I_{yy}}\rvert$	magnitude of second-order partial derivative in vertical direction
	$\lvert\mathbf{I_{xy}}\rvert$	magnitude of second-order mixed partial derivative
edge	$\sqrt{\mathbf{I}_\mathbf{x}^2 + \mathbf{I}_\mathbf{y}^2}$	magnitude of edge response
	$\tan^{-1}(\frac{\lvert\mathbf{I_y}\rvert}{\lvert\mathbf{I_x}\rvert})$	edge orientation
lab	l	luminance (LAB colour space)
	a	a channel (LAB colour space)
	b	b channel (LAB colour space)

III. Method

A. Covariance Descriptor

Let $\mathbf{x} \in \mathbb{R}^2$ denote the spatial coordinates of a pixel in an image Ω. Given a rectangular region of interest R in Ω and a feature mapping $\phi\colon \Omega \to \mathbb{R}^n$, the corresponding *region covariance* matrix is given by

$$\mathbf{\Lambda}_R = \frac{1}{|R|-1} \sum_{\mathbf{x}\in R} (\phi(\mathbf{x}) - \mu_R)(\phi(\mathbf{x}) - \mu_R)^\mathsf{T},$$

where $\mu_R = |R|^{-1} \sum_{\mathbf{x}\in R} \phi(\mathbf{x})$ and $|R|$ denotes the number of pixels in the region of interest. The matrix $\mathbf{\Lambda}_R$ describes the variations of the length-n feature vectors $\phi(\mathbf{x})$ as $\mathbf{x}$ varies over the region R. In this work we explore several candidate feature mappings obtained by selecting component elements from the set of features described in Table I.

B. Distance Measures

Covariance matrices are positive-definite and one can speak about a distance between a pair of covariance matrices once a distance measure is defined between members of the set of all real positive-definite matrices. Let $\mathit{Sym}(n)$ denote the set of all $n \times n$ symmetric real matrices, and let $\mathit{Sym}_+(n)$ denote the subset of $\mathit{Sym}(n)$ comprised of all $n \times n$ positive-definite matrices in $\mathit{Sym}(n)$. $\mathit{Sym}_+(n)$ can be endowed with a variety of distance measures [12], [17]. In what follows we shall consider three specific distances. One is the *Euclidean* metric given by

$$\mathrm{dist_E}(\mathbf{P}, \mathbf{Q}) = \|\mathbf{P} - \mathbf{Q}\|_\mathrm{F},$$

where $\|\cdot\|_\mathrm{F}$ denotes the Frobenius norm, and $\mathbf{P}$ and $\mathbf{Q}$ are members of $\mathit{Sym}_+(n)$. Another is the *Log-Euclidean* metric given by

$$\mathrm{dist_L}(\mathbf{P}, \mathbf{Q}) = \|\log \mathbf{P} - \log \mathbf{Q}\|_\mathrm{F},$$

where $\log$ denotes the principal matrix logarithm[1] [20]. Yet another distance measure is the *affine-invariant* metric given by

$$\mathrm{dist_A}(\mathbf{P}, \mathbf{Q}) = \left\|\log(\mathbf{P}^{-1}\mathbf{Q})\right\|_\mathrm{F} = \left\|\log\left(\mathbf{P}^{-1/2}\mathbf{Q}\mathbf{P}^{-1/2}\right)\right\|_\mathrm{F}$$

(cf. [21, Chap. XII]). The label "affine-invariant" reflects the fact that $\mathrm{dist_A}$ is invariant under each mapping of the form $\mathbf{P} \mapsto \mathbf{A}\mathbf{P}\mathbf{A}^\mathsf{T}$, where $\mathbf{A}$ is a real invertible matrix $\mathbf{A}$; that is,

$$\mathrm{dist_A}(\mathbf{P}, \mathbf{Q}) = \mathrm{dist_A}(\mathbf{A}\mathbf{P}\mathbf{A}^\mathsf{T}, \mathbf{A}\mathbf{Q}\mathbf{A}^\mathsf{T})$$

for all $\mathbf{P}, \mathbf{Q} \in \mathit{Sym}_+(n)$ and all invertible $n \times n$ matrices $\mathbf{A}$. The affine-invariant metric can alternatively be written as

$$\mathrm{dist_A}(\mathbf{P}, \mathbf{Q}) = \left(\sum_{i=1}^{n} \log^2 \lambda_i(\mathbf{P}^{-1}\mathbf{Q})\right)^{\frac{1}{2}}, \tag{1}$$

where $\lambda_i(\mathbf{P}^{-1}\mathbf{Q})$, $1 \le i \le n$, are the eigenvalues of $\mathbf{P}^{-1}\mathbf{Q}$. As the matrix $\mathbf{P}^{-1}\mathbf{Q}$ is similar to the symmetric matrix $\mathbf{P}^{-1/2}\mathbf{Q}\mathbf{P}^{-1/2}$, the eigenvalues $\lambda_i(\mathbf{P}^{-1}\mathbf{Q})$ are all positive and hence the right-hand side of (1) is well defined for all $\mathbf{P}$ and $\mathbf{Q}$ in $\mathit{Sym}_+(n)$.

Despite the differences in form, there is a unifying trait among the three selected distances—each distance can be interpreted as a geodesic distance with respect to a Riemannian metric on $\mathit{Sym}_+(n)$. To be more specific, note first that $\mathit{Sym}_+(n)$ forms an open cone of the $n(n+1)/2$-dimensional linear space $\mathit{Sym}(n)$. It then follows that $\mathit{Sym}_+(n)$ is a manifold whose tangent space at any foot point $\mathbf{P}$ can be identified with $\mathit{Sym}(n)$. Recall that a Riemannian metric on $\mathit{Sym}_+(n)$ is a family of inner products $\{g_\mathbf{P}\}_{\mathbf{P}\in \mathit{Sym}_+(n)}$ depending smoothly on the foot point $\mathbf{P}$. Given a Riemannian metric $\{g_\mathbf{P}\}_{\mathbf{P}\in \mathit{Sym}_+(n)}$, the length of a differentiable path $\gamma\colon [a, b] \to \mathit{Sym}_+(n)$ from $\mathbf{P} = \gamma(a)$ to $\mathbf{Q} = \gamma(b)$ is defined as

$$L(\gamma) = \int_a^b \sqrt{g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))}\, \mathrm{d}t.$$

The geodesic distance between points $\mathbf{P}$ and $\mathbf{Q}$ in $\mathit{Sym}_+(n)$ is given by

$$\inf\{L(\gamma) \mid \gamma \text{ is a differentiable path from } \mathbf{P} \text{ to } \mathbf{Q}\}.$$

Now, as it turns out, the distances $\mathrm{dist_E}$, $\mathrm{dist_L}$, and $\mathrm{dist_A}$ can be interpreted as geodesic distances corresponding to the Riemannian metrics

$$\begin{aligned} g^\mathrm{E}_\mathbf{P}(\mathbf{X}, \mathbf{Y}) &= \mathrm{Tr}(\mathbf{X}\mathbf{Y}), \\ g^\mathrm{L}_\mathbf{P}(\mathbf{X}, \mathbf{Y}) &= \mathrm{Tr}\left(\left(\mathrm{D}\log(\mathbf{P})[\mathbf{X}]\right)\left(\mathrm{D}\log(\mathbf{P})[\mathbf{Y}]\right)\right), \\ g^\mathrm{A}_\mathbf{P}(\mathbf{X}, \mathbf{Y}) &= \mathrm{Tr}(\mathbf{P}^{-1}\mathbf{X}\mathbf{P}^{-1}\mathbf{Y}) \\ & \qquad\qquad\qquad\qquad (\mathbf{X}, \mathbf{Y} \in \mathit{Sym}(n)), \end{aligned}$$

respectively. Here $\mathrm{D}\log(\mathbf{P})\colon \mathit{Sym}(n) \to \mathit{Sym}(n)$ denotes the Fréchet differential (or Fréchet derivative) of the $\log$ mapping at $\mathbf{P}$. There is no closed-form expression for $\mathrm{D}\log(\mathbf{P})$ amenable to easy implementation, however the following in-

[1] When $\mathbf{A}$ is an invertible matrix without non-negative eigenvalues, there exists a unique real logarithm of $\mathbf{A}$, called the principal logarithm and denoted $\log \mathbf{A}$, whose eigenvalues lie in the strip $\{z \in \mathbb{C} \mid -\pi < \mathrm{Im} z < \pi\}$ (cf. [18, Theorem 1.31], [19]).

tegral representation holds

$$\mathrm{D}\log(\mathbf{P})[\mathbf{X}] = \int_0^1 ((\mathbf{P}-\mathbf{I}_n)t+\mathbf{I}_n)^{-1}\mathbf{X}((\mathbf{P}-\mathbf{I}_n)t+\mathbf{I}_n)^{-1}\,\mathrm{d}t,$$

this being reminiscent of the numerical formula

$$\int_0^1 \frac{1}{((p-1)t+1)^2}\,\mathrm{d}t = -\frac{1}{p}\left[\frac{1}{(p-1)t+1}\right]_0^1 = \frac{1}{p} = (\log p)'$$

(cf. [22] [18, p. 272]).

The Euclidean and Log-Euclidean metrics are of direct significance from the machine learning perspective. It turns out that each of these metrics leads to a family $\{K_\lambda\}_{\lambda>0}$ of positive-definite kernels of the form

$$K_\lambda(\mathbf{P},\mathbf{Q}) = \exp\left(-\lambda\,\mathrm{dist}(\mathbf{P},\mathbf{Q})\right)$$

(cf. [12], [17], [23]). For any fixed $\lambda > 0$, the kernel K_λ can be used to construct a pattern classifier taking the form of a kernel support vector machine. In the case of the affine-invariant metric, there exist positive values λ for which K_λ is not positive definite [23], [24]. The significance of the affine-invariant metric stems mainly from its frequent appearance in various differential-geometric contexts including that of the Riemannian mean for positive-definite matrices [25].

C. *Singular Covariance Matrices*

In practice it is possible for the region covariance descriptor to occasionally produce a rank deficient covariance matrix. This can occur when some linear dependencies emerge in the feature vectors. For example, if the red, green, and blue channels are included in the feature vector, but the image happens to be greyscale, the colour channels will be linearly dependent and the corresponding covariance matrix will be rank deficient and no longer positive definite. To circumvent potential rank deficiency, we perform, given a covariance matrix, a rank revealing QR decomposition with column pivoting. The decomposition allows us to identify a subset of linearly independent features.

For a $n \times n$ covariance matrix $\mathbf{\Lambda}_R$, the QR factorisation of $\mathbf{\Lambda}_R$ with column pivoting is given by $\mathbf{\Lambda}_R\mathbf{P} = \mathbf{QR}$, where $\mathbf{P}$ is a permutation matrix, $\mathbf{Q}$ is orthogonal, and $\mathbf{R}$ is upper triangular with non-negative diagonal entries sorted in descending order of magnitude, all the matrices here being $n \times n$ matrices. The permutation matrix is chosen such that $\mathbf{R}$ has the structure

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix},$$

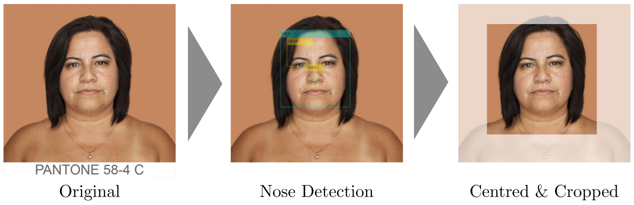
where $\mathbf{R}_{22}$ is a $k \times k$ matrix, $1 \leq k \leq n$, with k chosen so that $\|\mathbf{R}_{22}\|$ is small. The selection of k is motivated by the fact that $\sigma_{n-k+1}(\mathbf{\Lambda}_R) \leq \|\mathbf{R}_{22}\|$, where $\sigma_i(\mathbf{A})$ denotes the ith singular value of $\mathbf{A}$. The above inequality ensures that if $\|\mathbf{R}_{22}\|$ is small, then $\mathbf{\Lambda}_R$ has at least k small singular values, suggesting potential rank deficiency [26]. One can use the permutation matrix to identify which columns of $\mathbf{\Lambda}_R$ are involved in the formation of $\mathbf{R}_{22}$. Based on this information, the corresponding problematic features can then be removed from the feature vector thereby restoring linear independence and the requisite positive-definite property.

IV. Experiments

To evaluate the three distance measures we performed a number of comprehensive experiments using a novel face dataset. The purpose of the experiments was to investigate how each of the distance measures perform under varying image transformations and with varying feature sets.

A. *Dataset*

We evaluated the distance measures on images of human faces because faces strike a balance between similarity in the form of general structure, and dissimilarity in the form of age, sex, ethnicity, and personal individuality. Despite an abundance of established and publicly available face datasets [27]–[29], we utilised a new dataset called Humanæ which was created by the artist Angélica Dass[2]. This dataset offers numerous advantages for our targeted experiments. In particular, since it aims to capture the whole gamut of human skin tones, it spans broad age and ethnic groups. Moreover, all subjects are professionally photographed from a frontal perspective with controlled lighting and a uniform background. At the time of writing, the dataset contained 2,387 unique (no repeated persons) colour face images of dimension 500 × 500 pixels. We performed additional processing by centering all images on the nose and cropping to 319 × 319 pixels, thereby removing the text along the bottom of each image (see Figure 1).



Original

Nose Detection

Centred & Cropped

Fig. 1: Dataset processing.

B. *Comparison Types*

Our set of experiments can be split into two conceptual categories: *within* and *among*. Both rely on measuring distances between a set of images, the *comparable set*, and a single original image from the database, the *base image*.

1) within: In the *within* experiments the comparable set consisted solely of transformed versions of the base image, and the base image was compared to each transformed image using one of the distance measures. The purpose of these experiments was to characterise how the distance measures express—on average—the similarity between the base image and the transformed images.

2) among: In the *among* experiments the comparable set was composed of transformed versions of the base image as well as the entire dataset of non-transformed face images (the base image was excluded). The goal of these experiments was to determine whether the base image has greater similarity to transformed versions of itself rather than to other faces in the dataset. The experiments were posed as I query and retrieval

[2] All Humanæ images are copyrighted and used strictly with permission from Angélica Dass. Available at: http://humanae.tumblr.com/

tasks with I = 2,387. In particular, the nearest neighbours of every base image in the dataset were retrieved using each distance measure. To quantify the retrieval results for the ith base image ($i = 1, \ldots, I$), we first computed a precision per k nearest neighbours,

$$p_k^i = \frac{\text{number of retrieved transformed images}}{k}$$

for $k = 1, \ldots, K$, where K represents the total number of transformed base images. The precision scores were subsequently used to calculate the average precision for the ith query:

$$\mu_i = \left(\sum_{k=1}^{K} \mathbb{1}(k) \right)^{-1} \sum_{k=1}^{K} p_k^i \mathbb{1}(k),$$

where the indicator function $\mathbb{1}(k)$ is one if the kth nearest neighbour is a transformed base image and zero otherwise. Hence the term $\sum_{k=1}^{K} \mathbb{1}(k)$ counts the total number of transformed base images in a query of size K, and only precision values for which the kth nearest neighbour is also a correct retrieval contribute to the average precision score. In order to summarise how the distance measures perform we report the mean average precision given by

$$\text{MAP} = \frac{1}{I} \sum_{i=1}^{I} \mu_i.$$

If the problem at hand is recognising the same image or person under varying conditions or transformations, then a higher mean average precision value is desirable.

C. Feature Sets

In all experiments we considered various feature combinations in an attempt to identify combinations that perform well for particular images, transformations, or distance measures. The six specific different feature sets that we used are shown in Table I.

D. Transforms

We considered a broad range of geometric and photometric image transformations in order to increase the chance of discovering pertinent attributes for the distance measures.

1) Rotation (Geometric Transformation): Rotation transformations were obtained by rotating the base image anticlockwise by σ degrees, where $\sigma \in \{5, 10, 15, \ldots, 355\}$ (Figure 2). Images were post-transformation cropped to ensure that the pixels around the edges were from the original image and not black or hallucinated.

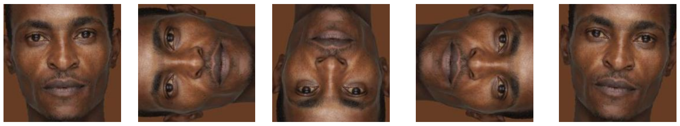


Fig. 2: The effect of the rotation transformation. From left to right $\sigma = 5, 90, 180, 270, 355$ degrees.

2) Gaussian Blur (Photometric Transformation): Gaussian blur transformations were applied with an isotropic Gaussian filter where the window size of the filer and the standard deviation of the Gaussian were both controlled by σ, with $\sigma \in \{2, 4, 6, \ldots, 100\}$ pixels (Figure 3).

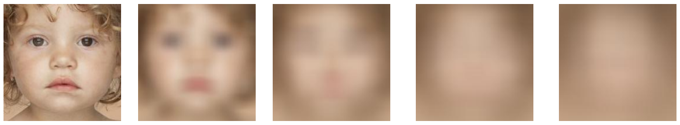


Fig. 3: The effect of the Gaussian blur transformation. From left to right $\sigma = 0, 25, 50, 75, 100$.

3) Gaussian Noise (Photometric Transformation): For the Gaussian noise transform each pixel was perturbed with independent Gaussian noise having mean 0.1 and variance $\sigma \in \{0.05, 0.1, 0.15, \ldots, 0.75\}$ (Figure 4).

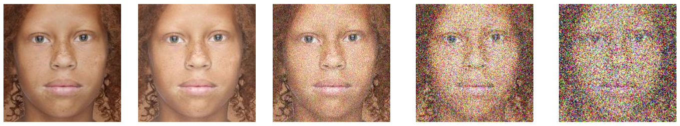


Fig. 4: The effect of the Gaussian noise transformation. From left to right $\sigma = 0, 0.15, 0.3, 0.45, 0.75$.

4) Brightness (Photometric Transformation): Brightness transformations were applied by transforming the RGB image to HSV and adding a value σ to the value channel, where $\sigma \in \{-1, -0.9, -0.8, \ldots, 0.9, 1\}$ (Figure 5).

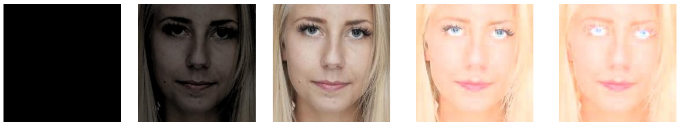


Fig. 5: The effect of the brightness (value) transformation. From left to right $\sigma = -1, -0.5, 0, 0.5, 1$.

5) Saturation (Photometric Transformation): The saturation transformation worked similarly to brightness: the image was converted from RGB to HSV, but now the saturation channel was modified, again by adding a value σ, where $\sigma \in \{-1, -0.9, -0.8, \ldots, 0.9, 1\}$ (Figure 6).

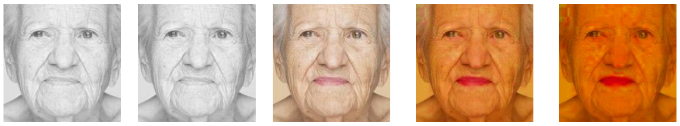


Fig. 6: The effect of the saturation transformation. From left to right $\sigma = -1, -0.5, 0, 0.5, 1$.

V. Results

The results for the *within* experiments are presented in the form of line graphs, with the x-axis denoting the value of the parameter governing a particular photometric or geometric image transformation and the y-axis depicting the value of the corresponding distance. Instead of displaying the raw distance values, we normalise the distance measures by ensuring that the area under the graph sums to one. This facilitates visual comparison between methods.

The ranking precisions gathered using the *among* experiments are shown using tables. Each row of the table represents one feature combination and the headings ‘Excl.’ and ‘Incl.’ serve as modifiers to indicate whether a particular feature combination was excluded or included in an experiment. For example, the row ‘None’ under the heading ‘Excl.’ signifies that all features were utilised and is considered our baseline. On the other hand, the row ‘rgb’ under the heading ‘Incl.’ indicates that only the rgb feature was utilised. Each cell value in the table represents the mean average precision, and the value in parentheses denotes the difference from the baseline. The numbers in the parentheses therefore quantify to what extent the inclusion or exclusion of a particular feature improves or decreases the retrieval performance when compared to the baseline.

A. Rotation

A visual inspection of the graph of the *within* rotation results presented in Figure 7 suggests that dist_E is biased for particular rotations because the graph is not symmetric with respect to a rotation of 180 degrees. This is especially evident when the graph of dist_E is compared with the graphs of dist_L and dist_A which exhibit symmetry. Overall dist_A performs best with rotations less than 90 or greater than 270 degrees resulting in the smallest distances. For both dist_L and dist_A a rotation of 180 degrees achieves a smaller distance measure than the rotations between 90 and 270 degrees.

The *among* rotation results (Table II) show that dist_A and dist_L perform substantially better than dist_E when including all features. The relevance of the colour features is also evident when considering the other two measures; the best result for each measure was achieved by only including colour features. This outcome is unsurprising, since the colour distribution in an image does not change substantially when the image is rotated.

TABLE II: Rotation Mean Average Precision

Excl.	dist_E	dist_L	dist_A
None	8.73	63.83	71.32
xy	9.85 (+1.12)	71.40 (+7.57)	84.43 (+13.11)
rgb	8.73 (-0.00)	48.78 (-15.05)	36.81 (**-34.51**)
∂	8.73 (0.00)	67.28 (+3.45)	73.85 (+2.53)
∂^2	8.73 (0.00)	66.33 (+2.51)	73.39 (+2.07)
$edge$	12.84 (**+4.10**)	65.56 (+1.73)	73.29 (+1.98)
lab	4.93 (**-3.81**)	56.31 (-7.52)	37.39 (**-33.93**)
Incl.	dist_E	dist_L	dist_A
rgb	41.46 (+32.73)	87.74 (+23.91)	89.99 (+18.67)
lab	53.15 (**+44.42**)	87.76 (+23.93)	88.42 (+17.10)
rgb, lab	53.15 (**+44.42**)	87.31 (+23.48)	**94.65** (+23.33)

B. Blur

The results of the *within* experiments (Figure 8) show that as the degree of blurring increases the resultant distance increases for all three measures at a reasonably linear rate.

The results of the *among* experiments (Table III) demonstrate the effectiveness of the dist_E measure, which achieves much higher precision than the other two metrics. Additionally, excluding the $edge$ features significantly improves results for dist_E, increasing the precision by 25%. Considering different active feature combinations, the best results are achieved by those which contain the colour (rgb and lab) and position (xy) features, with the xy and rgb combination achieving the best precision of 67%.

TABLE III: Gaussian Blur Mean Average Precision

Excl.	dist_E	dist_L	dist_A
None	13.07	6.01	7.33
xy	10.46 (-2.62)	5.96 (-0.05)	6.92 (-0.41)
rgb	13.07 (-0.00)	5.85 (-0.15)	6.05 (-1.28)
∂	13.07 (0.00)	6.09 (+0.08)	7.75 (+0.42)
∂^2	13.07 (0.00)	7.13 (+1.12)	9.99 (**+2.66**)
$edge$	37.96 (**+24.89**)	6.05 (+0.04)	7.59 (+0.25)
lab	7.28 (-5.79)	5.78 (-0.23)	6.06 (-1.28)
Incl.	dist_E	dist_L	dist_A
xy, rgb	67.23 (**+54.16**)	10.04 (+4.03)	16.51 (**+9.17**)
xy, lab	37.97 (+24.89)	12.70 (**+6.69**)	16.02 (+8.68)
xy, rgb, lab	37.96 (+24.89)	8.00 (+1.99)	12.63 (+5.30)

C. Noise

The *within* results (Figure 9) show a fairly steady increase in distance as the level of noise increases. However dist_E and dist_A grow somewhat slower for small noise values compared to dist_L.

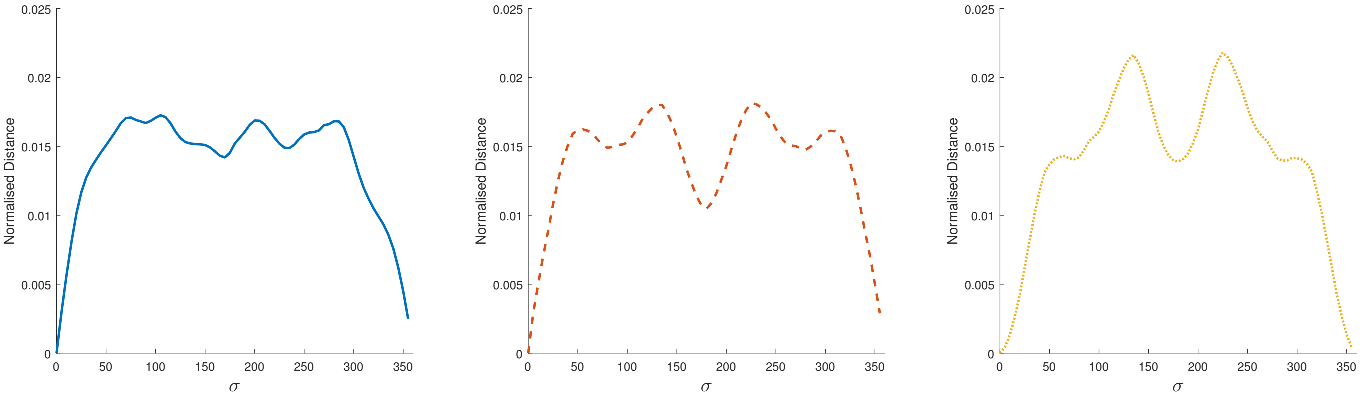
The *among* tests (Table IV) show that dist_E performs best overall, especially when the xy features are included. The highest precision is achieved with the xy and rgb combination. For measure dist_A and dist_L precision can be improved by excluding the colour features rgb or lab whilst keeping all other features. Alternatively, including only xy and one of the colour features also leads to similar precision improvements. For dist_A the xy and ∂^2 features also increase precision substantially, and for dist_L the xy and ∂ features have the opposite effect, decreasing the precision.

TABLE IV: Gaussian Noise Mean Average Precision

Excl.	dist_E	dist_L	dist_A
None	30.19	19.22	13.09
xy	22.83 (-7.36)	17.69 (-1.53)	12.89 (-0.20)
rgb	30.19 (-0.00)	26.66 (**+7.44**)	30.42 (**+17.33**)
∂	30.19 (0.00)	18.52 (-0.70)	13.03 (-0.06)
∂^2	30.19 (0.00)	18.00 (-1.22)	12.97 (-0.12)
$edge$	43.90 (**+13.71**)	18.78 (-0.44)	13.04 (-0.05)
lab	21.89 (-8.30)	24.85 (**+5.63**)	30.12 (**+17.03**)
Incl.	dist_E	dist_L	dist_A
xy, rgb	**78.38** (**+48.19**)	24.04 (**+4.82**)	27.91 (**+14.83**)
xy, lab	43.90 (+13.71)	26.34 (**+7.12**)	28.49 (**+15.40**)
xy, rgb, lab	43.90 (+13.71)	17.12 (-2.10)	12.87 (-0.22)
xy, ∂	48.48 (+18.29)	12.60 (-6.62)	12.99 (-0.10)
xy, ∂^2	44.27 (+14.08)	12.83 (-6.40)	14.35 (+1.27)

D. Brightness

The *within* results (Figure 10) show that all distance measure are affected more significantly when brightness is



(a) $\mathrm{dist}_{\mathrm{E}}$

(b) $\mathrm{dist}_{\mathrm{L}}$

(c) $\mathrm{dist}_{\mathrm{A}}$

Fig. 7: Effect of rotation on distances (normalised). All features included. Averaged over entire dataset.

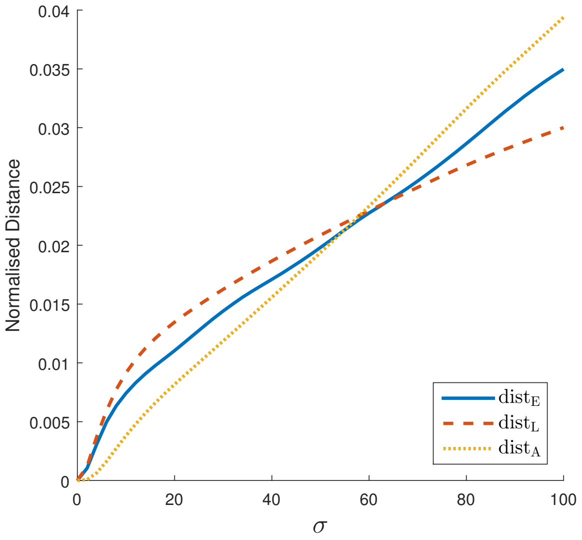


Fig. 8: Effect of Gaussian blur on distances (normalised). All features included. Averaged over entire dataset.

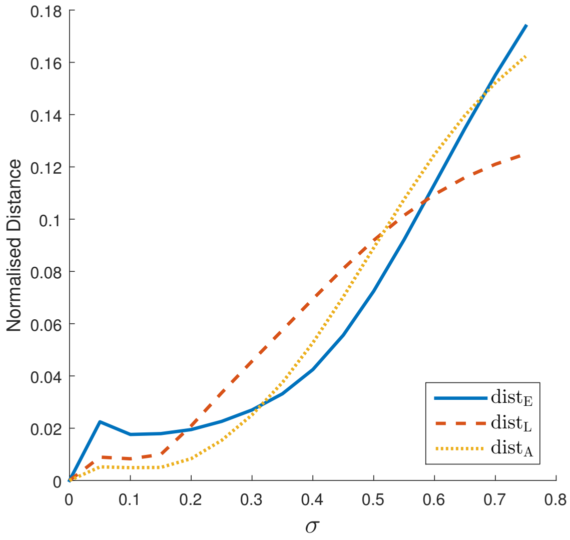


Fig. 9: Effect of Gaussian noise on distances (normalised). All features included. Averaged over entire dataset.

decreased rather than when it is increased. In the case of $\mathrm{dist}_{\mathrm{L}}$ the distances grow at similar rate regardless of whether brightness is increased or decreased moderately, whereas for $\mathrm{dist}_{\mathrm{E}}$ and $\mathrm{dist}_{\mathrm{A}}$ the distances grow at a higher rate when brightness is decreased.

The *among* results (Table V) show that $\mathrm{dist}_{\mathrm{E}}$ is the most effective measure; however, this effectiveness is greatly decreased with the exclusion of the xy features. The exclusion of the $edge$ features improves the effectiveness of $\mathrm{dist}_{\mathrm{E}}$. For the $\mathrm{dist}_{\mathrm{A}}$ measure in particular, excluding the colour features rgb or lab results in better precision. Excluding everything but one of the colour features and the xy features increase performance to a similar extent. This result suggests a counter-action between colour and edge features. The best performing measure is $\mathrm{dist}_{\mathrm{E}}$ with the features xy and lab included.

E. Saturation

The *within* results (Figure 11) show that for distance measures $\mathrm{dist}_{\mathrm{L}}$ and $\mathrm{dist}_{\mathrm{A}}$ increasing the saturation has a lesser effect on the distance than decreasing it, whereas the opposite is true for the $\mathrm{dist}_{\mathrm{E}}$ measure. In the case of diminishing the saturation for $\mathrm{dist}_{\mathrm{E}}$, once it is desaturated past approximately 0.2 the distance remains relatively constant. For $\mathrm{dist}_{\mathrm{L}}$ and $\mathrm{dist}_{\mathrm{A}}$ the behaviour that occurs when desaturating the image is not standard across all images, with the distance continuously increasing for some while for others the distance starts decreasing again past a particular threshold. This instability can be attributed to rank deficiencies described in Section III-C and the fact that the QR decomposition with column pivoting is applied in order to remove linear dependencies.

The *among* results (Table VI) show that overall $\mathrm{dist}_{\mathrm{E}}$

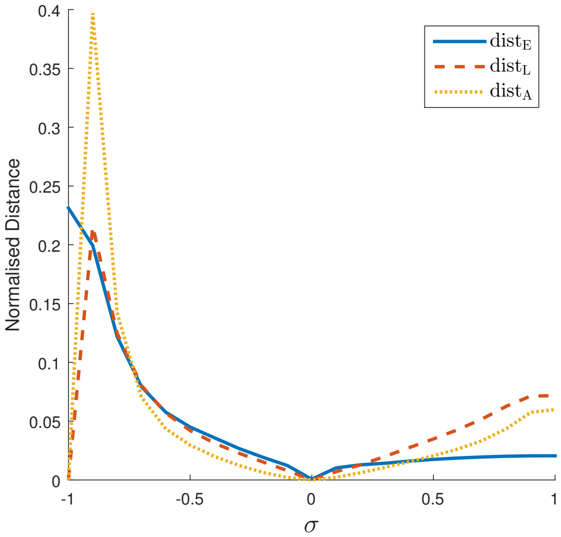


Fig. 10: Effect of value on distances (normalised). All features included. Averaged over entire dataset.

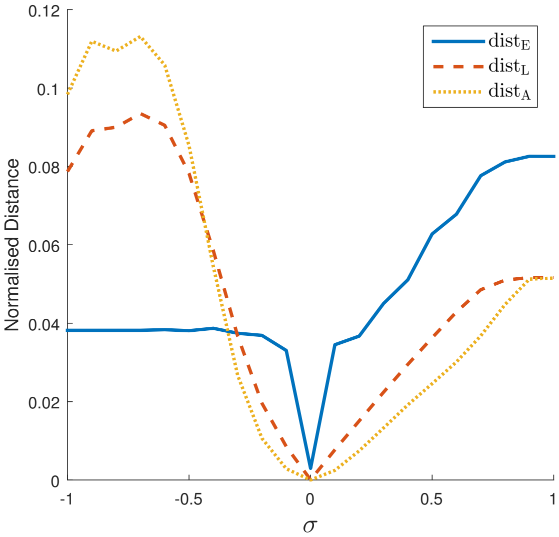


Fig. 11: Effect of saturation on distances (normalised). All features included. Averaged over entire dataset.

TABLE V: Brightness Mean Average Precision

Excl.	$\mathrm{dist_E}$	$\mathrm{dist_L}$	$\mathrm{dist_A}$
None	27.13	23.16	19.67
xy	19.28 (**-7.85**)	22.28 (-0.88)	19.18 (-0.49)
rgb	27.13 (-0.00)	24.97 (+1.82)	27.63 (**+7.95**)
∂	27.13 (0.00)	22.84 (-0.32)	19.54 (-0.14)
∂^2	27.13 (0.00)	22.57 (-0.59)	19.41 (-0.26)
$edge$	34.76 (**+7.62**)	22.99 (-0.17)	19.57 (-0.10)
lab	22.09 (**-5.05**)	22.79 (-0.37)	27.64 (**+7.97**)
rgb, lab	22.05 (**-5.08**)	18.86 (**-4.30**)	20.83 (+1.15)
Incl.	$\mathrm{dist_E}$	$\mathrm{dist_L}$	$\mathrm{dist_A}$
xy, rgb	34.02 (+6.88)	19.68 (-3.47)	26.04 (**+6.37**)
xy, lab	**34.76** (**+7.62**)	24.34 (+1.18)	26.25 (**+6.58**)
xy, rgb, lab	**34.76** (**+7.62**)	22.03 (-1.13)	19.19 (-0.49)
xy, ∂	32.15 (+5.02)	18.52 (-4.64)	18.65 (-1.03)
xy, ∂^2	34.40 (+7.27)	18.56 (-4.59)	19.16 (-0.51)

TABLE VI: Saturation Mean Average Precision

Excl.	$\mathrm{dist_E}$	$\mathrm{dist_L}$	$\mathrm{dist_A}$
None	47.07	14.67	14.56
xy	23.52 (**-23.56**)	14.41 (-0.26)	14.23 (-0.33)
rgb	47.07 (-0.00)	19.06 (**+4.39**)	21.21 (**+6.65**)
∂	47.07 (0.00)	14.50 (-0.17)	14.49 (-0.07)
∂^2	47.07 (0.00)	14.29 (-0.37)	14.18 (-0.38)
$edge$	37.75 (-9.32)	14.57 (-0.09)	14.72 (+0.16)
lab	40.23 (-6.85)	19.91 (**+5.24**)	23.26 (**+8.70**)
Incl.	$\mathrm{dist_E}$	$\mathrm{dist_L}$	$\mathrm{dist_A}$
xy, ∂	70.49 (+23.42)	14.09 (-0.58)	16.04 (+1.48)
xy, ∂^2	78.00 (+30.93)	16.27 (+1.60)	26.77 (**+12.21**)
xy, ∂, ∂^2	**81.39** (**+34.32**)	19.20 (+4.53)	29.33 (**+14.77**)
$\partial, edge$	16.07 (**-31.00**)	14.68 (+0.01)	17.15 (+2.59)
$\partial^2, edge$	16.06 (**-31.01**)	17.70 (+3.04)	20.77 (+6.20)
$\partial, \partial^2, edge$	16.09 (**-30.98**)	20.16 (+5.49)	24.87 (**+10.30**)

outperforms the other two measures, and that excluding the xy features greatly decreases the precision. Specifically, measure $\mathrm{dist_E}$ with the xy, ∂ and ∂^2 achieves the best performance. Excluding the colour features rgb or lab while maintaining the other features increases the precision for measures $\mathrm{dist_L}$ and $\mathrm{dist_A}$. These results suggest the importance of position and edge based features for saturation changes.

VI. Discussion

As a whole the results paint a rather complex picture—there is no distance measure which works best in all situations. Moreover, the inclusion or exclusion of a single feature can have a dramatic impact on the efficacy of a distance measure. Hence the selection of a suitable collection of features for a particular problem must be guided by extensive empirical analysis.

A surprising outcome of the experiments was the excellent retrieval performance observed for the $\mathrm{dist_E}$ measure for Gaussian noise and blur transformations when the position feature (xy) was combined with a colour feature (rgb or lab). This particular finding warrants further investigation, since it suggests that the region covariance descriptor may be useful for methods that perform image super-resolution, deblurring, and denoising based on matching and retrieval of image patches [30]. The region covariance descriptor could also be beneficial for algorithms that perform image restoration under the assumption that an image patch can be encoded as a sparse linear combination of basis images. Such paradigms typically involve a learning step wherein a suitable collection of basis images is constructed from numerous example images. The assessment of similarity between image patches is of essential importance in the learning process [31]. Our experiments suggest that the region covariance descriptor can be used to capture an appropriate notion of similarity even under considerable noise or image blur.

VII. Conclusion

Our work has explored various aspects of the region covariance descriptor. We discussed three different distance measures that are frequently utilised and explained their significance. We also explored the efficacy of the distance measures through extensive targeted experiments in which we investigated numerous feature combinations. Our findings suggest that no specific distance measure is best for all scenarios, and that the choice of features can have a dramatic impact on performance. In future work we intend to: (1) replicate the classification results reported in [17] and [32], and (2) conduct a thorough analysis of the datasets in those papers in order to understand why certain distance measures work best in a particular context. We also intend to investigate the potential utility of the region covariance descriptor for various image restoration tasks.

Acknowledgements

This research was supported by the Australian Research Council. We would like to thank Angélica Dass for granting us permission to use her images.

References

[1] A. Rosenfeld and G. J. van der Brug, “Coarse-fine template matching,” *IEEE Trans. Syst., Man, Cybern.*, vol. 7, no. 2, pp. 104–107, 1977.

[2] C. Undurraga and D. Mery, “Improving tracking algorithms using saliency,” in *Proc. 16th Iberoamerican Congress on Pattern Recognition*, ser. Lecture Notes in Computer Science, 2011, vol. 7042, pp. 141–148.

[3] A. Alahi, M. Bierlaire, and M. Kunt, “Object detection and matching with mobile cameras collaborating with fixed cameras,” in *Proc. ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.

[4] A. Alahi, P. Vandergheynst, M. Bierlaire, and M. Kunt, “Cascade of descriptors to detect and track objects across any network of cameras,” *Comput. Vis. Image Underst.*, vol. 114, no. 6, pp. 624–640, 2010.

[5] O. Tuzel, F. M. Porikli, and P. Meer, “Region covariance: a fast descriptor for detection and classification,” in *Proc. 9th European Conf. Computer Vision*, ser. Lecture Notes in Computer Science, vol. 3952, 2006, pp. 589–600.

[6] J. Yao and J. M. Odobez, “Fast human detection from videos using covariance features,” in *Proc. 8th ECCV Workshop on Visual Surveillance*, 2008.

[7] ——, “Fast human detection from joint appearance and foreground feature subset covariances,” *Comput. Vis. Image Underst.*, vol. 115, no. 10, pp. 1414–1426, 2011.

[8] W. Ayedi, H. Snoussi, and M. Abid, “A fast multi-scale covariance descriptor for object re-identification,” *Pattern Recognition Lett.*, vol. 33, no. 14, pp. 1902–1907, 2012.

[9] E. Erdem and A. Erdem, “Visual saliency estimation by nonlinearly integrating features using region covariances,” *J. Vision*, vol. 13, no. 4, pp. 1–20, 2013.

[10] O. Tuzel, F. Porikli, and P. Meer, “Pedestrian detection via classification on Riemannian manifolds,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1713–1727, 2008.

[11] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, “Combining multiple manifold-valued descriptors for improved object recognition,” in *Proc. Int. Conf. Digital Image Computing: Techniques and Applications*, 2013, pp. 1–6.

[12] ——, “Kernel methods on the Riemannian manifold of symmetric positive definite matrices,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 2013, pp. 73–80.

[13] L. Qin, H. Snoussi, and F. Abdallah, “Adaptive covariance matrix for object region representation,” in *Proc. 5th Int. Conf. on Digital Image Processing*, ser. Proc. SPIE, 2013, vol. 8878, pp. 887 848–1–887 848–7.

[14] V. Sulic, J. Perš, M. Kristan, and S. Kovacic, “Histogram of oriented gradients and region covariance descriptor in hierarchical feature-distribution scheme,” in *Proc. 19th Int. Electrotechnical and Computer Science Conf.*, 2010, pp. 229–232. [Online]. Available: http://vision.fe.uni-lj.si/docs/danas/SulicERK2010FINAL.pdf

[15] P. C. Cargill, C. U. Rius, D. M. Quiroz, and A. Soto, “Performance evaluation of the covariance descriptor for target detection,” in *Proc. Int. Conf. of the Chilean Computer Science Society*, 2009, pp. 133–141.

[16] M. Harandi, M. Salzmann, and F. Porikli, “Bregman divergences for infinite dimensional covariance matrices,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 1003–1010.

[17] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, “Kernel methods on the Riemannian manifolds with Gaussian RBF kernels,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015.

[18] N. J. Higham, *Functions of Matrices: Theory and Computation*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2008.

[19] W. J. Culver, “On the existence and uniqueness of the real logarithm of a matrix,” *Proc. Amer. Math. Soc.*, vol. 17, pp. 1146–1151, 1966.

[20] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, “Fast and simple calculus on tensors in the Log-Euclidean framework,” in *Proc. 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, ser. Lecture Notes in Computer Science, 2005, vol. 3749, pp. 115–122.

[21] S. Lang, *Fundamentals of Differential Geometry*. New York: Springer, 1999.

[22] L. Dieci, B. Morini, and A. Papini, “Computational techniques for real logarithms of matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 3, pp. 570–593, 1996.

[23] A. Feragen, F. Lauze, and S. Hauberg, “Geodesic exponential kernels: when curvature and linearity conflict,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 3032–3042.

[24] M. T. Harandi, C. Sanderson, A. Wiliem, and B. C. Lovell, “Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures,” in *Proc. IEEE Workshop on Applications of Computer Vision*, 2012, pp. 433–439.

[25] R. Bhatia, “The Riemannian mean of positive matrices,” in *Matrix Information Geometry*, F. Nielsen and R. Bhatia, Eds. Heidelberg: Springer, 2013, pp. 35–51.

[26] Y. P. Hong and C.-T. Pan, “Rank-revealing QR factorizations and the singular value decomposition,” *Mathematics of Computation*, vol. 58, no. 197, pp. 213–232, 1992.

[27] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” University of Massachusetts, Amherst, Tech. Rep. 07-49, 2007.

[28] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, “Attribute and simile classifiers for face verification,” in *Proc. 12th Int. Conf. Computer Vision*, 2009, pp. 365–372.

[29] H.-W. Ng and S. Winkler, “A data-driven approach to cleaning large face datasets,” in *Proc. IEEE Int. Conf. on Image Processing*, 2014, pp. 343–347.

[30] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising with block matching and 3D filtering,” in *Proc. SPIE Electronic Imaging: Algorithms and Systems, Neural Networks, and Machine Learning*, vol. 6064, 2006, pp. 606 414–1–606 141–12.

[31] J. C. Ferreira, E. Vural, and C. Guillemot, “Geometry-aware neighborhood search for learning local models for image reconstruction,” [arXiv preprint arXiv:1505.01429, May 2015].

[32] M. Harandi and M. Salzmann, “Riemannian coding and dictionary learning: kernels to the rescue,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 3926–3935.

[33] *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015.