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

† School of Engineering
University of KwaZulu-Natal, Durban 4041, South Africa
INTRODUCTION

A modern computer vision pipeline for generic image classification and recognition consists of three broad conceptual steps:

• selecting suitable image descriptors
• defining a measure of similarity between feature descriptors
• learning a classification rule that uses the feature descriptors and corresponding similarity measure to determine what the image represents
INTRODUCTION

A modern computer vision pipeline for generic image classification and recognition consists of three broad conceptual steps:

• selecting suitable image descriptors
• defining a measure of similarity between feature descriptors
• learning a classification rule that uses the feature descriptors and corresponding similarity measure to determine what the image represents
INTRODUCTION

A modern computer vision pipeline for generic image classification and recognition consists of three broad conceptual steps:

• selecting suitable image descriptors
• defining a measure of similarity between feature descriptors
• learning a classification rule that uses the feature descriptors and corresponding similarity measure to determine what the image represents
INTRODUCTION

A modern computer vision pipeline for generic image classification and recognition consists of three broad conceptual steps:

• selecting suitable image descriptors
  – region covariance descriptor

• the definition of a measure of similarity between feature descriptors
  – distance between covariance matrices
A modern computer vision pipeline for generic image classification and recognition consists of three broad conceptual steps:

• selecting suitable image descriptors
  – region covariance descriptor
• the definition of a measure of similarity between feature descriptors
  – distance between covariance matrices
MOTIVATION

• Region covariance descriptor has proven to be useful in numerous computer vision applications.
• The properties of the descriptor are not well understood or documented.
REGION COVARIANCE DESCRIPTOR

$\Omega$ image

$x$ spatial coordinates of a pixel in image

$R$ rectangular region of interest in image

$\phi: \Omega \to \mathbb{R}^n$ mapping from pixels to length-n feature vectors

$\Lambda_R$ n-by-n covariance matrix

\[
\Lambda_R = \frac{1}{|R| - 1} \sum_{x \in R} (\phi(x) - \mu_R)(\phi(x) - \mu_R)^T
\]

\[
\mu_R = \frac{1}{|R|} \sum_{x \in R} \phi(x)
\]

$|R|$ number of pixels in R

mean feature
FEATURE MAPPINGS

- spatial x coordinate
- spatial y coordinate
- red channel
- green channel
- blue channel
- magnitude of first-order partial derivative in horizontal direction
- magnitude of first-order partial derivative in vertical direction
- magnitude of second-order partial derivative in horizontal direction
- magnitude of second-order partial derivative in vertical direction
- magnitude of second-order mixed partial derivative
- magnitude of edge response
- edge orientation
- luminance (LAB colour space)
- a channel (LAB colour space)
- b channel (LAB colour space)
How should one define $\text{dist}(\Lambda_R, \Lambda_R)$?
DISTANCE BETWEEN COVARIANCE MATRICES

$\text{Sym}(n)$  
set of all $n \times n$ symmetric real matrices

$\text{Sym}_+(n)$  
subset of positive definite matrices in $\text{Sym}(n)$

$P, Q$  
covariance matrices in $\text{Sym}_+(n)$

$\| \cdot \|_F$  
Frobenius norm

\[
\begin{align*}
\text{dist}_E(P, Q) &= \|P - Q\|_F & \text{Euclidean metric} \\
\text{dist}_L(P, Q) &= \|\log P - \log Q\|_F & \text{Log-Euclidean metric} \\
\text{dist}_A(P, Q) &= \|\log (P^{-1}Q)\|_F \\
&= \|\log (P^{-1/2}QP^{-1/2})\|_F & \text{Affine-invariant metric}
\end{align*}
\]
How do features and distance measures influence the similarity between two images?
DATASET

- Diverse images of human faces $500 \times 500$ pixels
- Processing by centering all images on the nose and cropping to $319 \times 319$ pixels
TRANSFORMATIONS

saturation

brightness

blur

noise

rotation
EXPERIMENTS

• within: comparable set $\triangleq$ transformed base images
EXPERIMENTS (Cont.)

- among: comparable set ≜ transformed base images + entire dataset
RESULTS: Different Base Image

Features: $x, y, r, g, b$
Distance: Euclidean
Transform: Blur

base image

ascending order
distance to base image

Features: Same
Distance: Same
Transform: Same
RESULTS: Different Feature Set

Features: \( x, y, r, g, b, l, a, b \)
Distance: Euclidean
Transform: Blur

Features: \( x, y, r, g, b, \sqrt{l_x^2 + l_y^2}, \tan^{-1}\left(\frac{|l_y|}{|l_x|}\right), l, a, b \)
Distance: Same
Transform: Same
RESULTS: Different Distance

Features: x, y, r, g, b
Distance: Euclidean
Transform: Blur

Features: Same
Distance: Log-Euclidean
Transform: Same
RESULTS: Different Distance

Features: $x, y, r, g, b$
Distance: Euclidean
Transform: Blur

Features: Same
Distance: Affine Invariant
Transform: Same
RESULTS: Different Transform

Features: $x, y, r, g, b$
Distance: Euclidean
Transform: Blur

Features: Same
Distance: Same
Transform: Rotation
RESULTS: Different Distance for Different Problems

Features: $|I_{xx}|, |I_{yy}|, |I_{xy}|, \sqrt{I_x^2 + I_y^2}, \tan^{-1}\left(\frac{I_y}{I_x}\right)$
Distance: Euclidean
Transform: Rotation

Features: Same
Distance: Affine Invariant
Transform: Same
• No distance measure works best in all situations.
• Inclusion or exclusion of a single feature can have a dramatic impact.
• Selection of features must be guided by extensive empirical analysis.
• Excellent retrieval performance observed for the $dist_E$ measure for Gaussian noise and blur transformations when the position feature ($xy$) was combined with a colour feature ($rgb$ or $lab$).
DISCUSSION

• No distance measure works best in all situations.
• Inclusion or exclusion of a single feature can have a dramatic impact.
• Selection of features must be guided by extensive empirical analysis.
• Excellent retrieval performance observed for the \( \text{dist}_E \) measure for Gaussian noise and blur transformations when the position feature \((xy)\) was combined with a colour feature \((rgb \text{ or } lab)\).
DISCUSSION

• No distance measure works best in all situations.
• Inclusion or exclusion of a single feature can have a dramatic impact.
• Selection of features must be guided by extensive empirical analysis.
• Excellent retrieval performance observed for the $dist_E$ measure for Gaussian noise and blur transformations when the position feature ($xy$) was combined with a colour feature ($rgb$ or $lab$).
DISCUSSION

• No distance measure works best in all situations.
• Inclusion or exclusion of a single feature can have a dramatic impact.
• Selection of features must be guided by extensive empirical analysis.
• Excellent retrieval performance observed for the $dist_{E}$ measure for Gaussian noise and blur transformations when the position feature ($xy$) was combined with a colour feature ($rgb$ or $lab$).
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.
QUESTIONS