Computer Vision and Image Understanding 000 (2017) 1–11

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

# Naturally combined shape-color moment invariants under affine transformations☆

Ming Gong[a,b,c,1], You Hao[a,b,*], Hanlin Mo[a,b], Hua Li[a,b]

[a] Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[b] University of Chinese Academy of Sciences, Beijing, China
[c] AI & Research Group, Microsoft Search Technology Center Asia, Beijing, China

A R T I C L E I N F O

*Article history:*
Received 26 October 2016
Revised 31 May 2017
Accepted 20 July 2017
Available online xxx

*Keywords:*
Invariants
Color
Shape
Descriptor
Dual-affine transformations

A B S T R A C T

We proposed a kind of naturally combined shape-color affine moment invariants (SCAMI), which consider both shape and color affine transformations simultaneously in one single system. In the real scene, color and shape deformations always exist in images simultaneously. Simple shape invariants or color invariants cannot be qualified for this situation. The conventional method is just to make a simple linear combination of the two factors. Meanwhile, the manual selection of weights is a complex issue. Our construction method is based on the multiple integration framework. The integral kernel is assigned as the continued product of the shape and color invariant cores. It is the first time to directly derive an invariant to dual affine transformations of shape and color. The manual selection of weights is no longer necessary, and both the shape and color transformations are extended to affine transformation group. With the various of invariant cores, a set of lower-order invariants are constructed and the completeness and independence are discussed detailed. A set of SCAMIs, which called SCAMI24, are recommended, and the effectiveness and robustness have been evaluated on both synthetic and real datasets.



## 1. Introduction

With the development of information technology, the number of images in people's daily life is increasing rapidly. In computer vision, the extraction of image description has been one of the most important tasks. Shape and color are two basic types of information to people's visual cognition and play very important roles in image analysis and understanding. Shape features reflect the position, size, and shape information and color features reflect different spectral reflecting attributes of the object's surface, as shown in Fig. 1.

The real imaging environment is intricate, which is influenced by different scene illumination, camera sensors, and the reflective characteristics of the objects. Images captured from real scene are always degraded, hence their color and shape are not consistent. This means the color of images are different, and the geometric deformation like scaling, rotation, and skewing are occurring. Various approaches have been proposed to recognize images of the same objects under such geometric and photometric deformations. An effective way is to extract invariant features.

The 2-D geometric moment invariants were firstly proposed by Hu in 1962 for character recognition (Hu, 1962), which are invariant to similarity transformation. The seven invariants can describe some of features of shape and played important roles in pattern recognition. In Flusser and Suk (1993), Suk and Flusser extend the moment invariants from similarity transformation to affine transformation. They are invariant under the affine transformation. The main advantage of invariant features is their invariance under given transformations and there is no need to consider the corresponding deformations in the imaging process. Hence, it was even argued that object recognition is the search for invariants (Alferez, 1993).

Color information is useful for object recognition. Many approaches extracting color information are based on color histograms (Funt and Finlayson, 1995; Healey and Slater, 1994) and color moment (Stricker and Orengo, 1995). They make full use of color information but have no color consistency, so are sensitive to illumination changes. Some other color constant descriptors in

☆ This work is supported by the National Science Foundation of China under Grant 61379082 and 61227802.

* Corresponding author at: Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, No.6 Academy of Science South Road, Haidian District Beijing, China:

*E-mail addresses:* yiming10.13@163.com (M. Gong), haoyou@ict.ac.cn (Y. Hao), mohanlin@ict.ac.cn (H. Mo), lihua@ict.ac.cn (H. Li).

[1] This work is mainly done when Ming Gong pursued her doctor degree in the Institute of Computing Technology. She is now working in Microsoft AI & Research Group.

http://dx.doi.org/10.1016/j.cviu.2017.07.003
1077-3142/

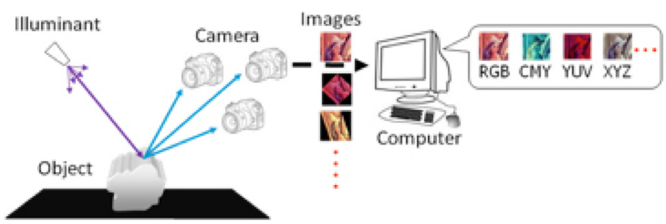


**Fig. 1.** The factors influencing the results during imaging procedure: illumination, camera sensors, and the surface reflective characteristics.

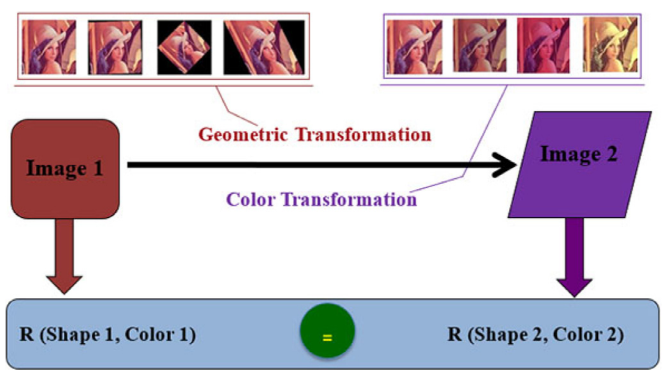


**Fig. 2.** The analysis and construction method of shape and color invariant features.

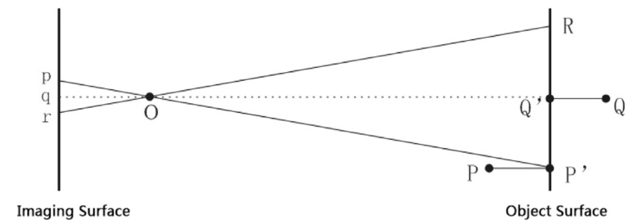


**Fig. 3.** The model of projection approximated by affine transformation.

Li et al. (2009), Gevers and Smeulders (1999) can deal with the degeneration of illumination. Gong et al. (2013) proposed a kind of color affine moment invariants which are applicable to more complicated color variance and robust to shape deformation to certain extent. The limitation of the color descriptor is that they do not exploit any spatial information of the object, which leads to vital information lost.

Lots of efforts have been made on the invariant features which can deal with both shape and color deformations. In common practices, shape and color descriptors are extracted independently, as shown in Fig. 2. Consequently, they are not very robust in the real condition. It is difficult to combine those two types of information in one descriptor effectively. The conventional method is just to make a simple linear combination of the two factors, but manual intervention for the selection of weights is a complex issue. Wang et al. proposed a kind of modified Hu invariant moments Wang et al. (2008), which contains both color information and shape information. However, this method is just applicable for the gray-level degradation. Mindru et al. proposed a kind of generalized color moment invariants (Mindru et al., 2004), extending the moment invariants obtained by Lie group methods detailed in Moons et al. (1995), Van Gool et al. (1995), which considers both shape and color deformations. The generalized color moment invariants do not change under geometric affine and photometric diagonal transformations, which get good performance on the experimental datasets. For the diagonal transformation of the color with 3-bands (R,G,B), transformation parameters of three color bands can be separated easily, hence the invariants can be constructed in an arbitrary combination of color bands. However, for the affine transformation, it is impossible to separate parameters by different color bands that this method cannot be extended to the color affine transformation.

Many works apply invariants into shape's contours to extract shape invariant features (Mundy and Zisserman, 1992; Mundy et al., 1994). The limitation of this kind method is that the object contours should be extracted robustly, which is difficult to reach in real scene. Alferez and Wang proposed a method considering both geometric and illumination deformations (Alterez and Wang, 1999). These invariants are based on the sampling curves extracted from the image, which can be the contour of imaged objects or some characteristic curves. This method gets good performance on experimental dataset. The limitation is that the objects' contours should be properly extracted in advance and the weighted average between geometry and illumination invariants also needs manual intervention (Alterez and Wang, 1999).

This paper is dedicated to research on Shape-Color Affine Moment Invariants (SCAMI). The construction method is based on the multiple integration framework, which can be extended easily along with the various integral kernels. The integral kernel is assigned as the continued product of the shape and color invariant cores, which is based on the geometric invariant primitives.

The main advantage of SCAMI is that they naturally and intrinsically unify shape and color factors together, which is the first time to directly derive an invariant to shape-color dual-affine transformations. Furthermore, the manual selection of weights is no longer necessary. In addition, they can be extended to higher order and dimension easily.

The paper is organized as follows. In Section 2, the background to construct SCAMI is presented, including the geometric model and the illumination model. In Section 3, the invariant constructing framework will be claimed and the derivation of the SCAMI is conducted. Several cases are listed in Section 4 and the independent of them are tested. In Section 5, the experiment results and conclusion are listed.

## 2. Background

The Shape-Color Affine Moment Invariants (SCAMI) are invariant under the shape affine transformation and color affine transformation. In the real imaging process, the imaging system and conditions are uncertain, which is influenced by different scene illumination, camera sensors, and many other factors. Images captured from real scene are always degraded, hence their color and shape are not consistent. It means that color and shape degradation will bring into the image because of the complex imaging conditions and different systems. In order to deal with these degradations, the transformation mode should be established.

### 2.1. Shape transformation

Similar to the imaging mechanism of human's eyes, the camera model is projective transformation from 3D to 2D. When the camera takes images of the same object in Healey and Slater (1994) different viewpoints, the relationship of the images follows the projective transformation.

Fortunately, in the imaging process, when the distance between the camera and the object is far enough, the projective transformation can be well approximated by affine transformation, as shown in the Fig. 3.

The transformation from $(P, Q)$ to $(p, q)$ is projective transformation. And the similar transformation from [4] $(P', Q')$ to $(p, q)$ is affine transformation. Therefore, we take affine geometric transformation into consideration. Assume spatial coordinates $(x, y)$ in

Please cite this article as: M. Gong et al., Naturally combined shape-color moment invariants under affine transformations, Computer Vision and Image Understanding (2017), http://dx.doi.org/10.1016/j.cviu.2017.07.003