INVARIANT REPRESENTATION IN IMAGE PROCESSING

Arnold W. M. Smeulders and Jan-Mark Geusebroek and Theo Gevers

ISIS group, Faculty of Science, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
{ smeulders, geusebroek, gevers }@science.uva.nl

ABSTRACT

The paper discusses the role of invariance in image processing, specifically the desire to discriminate against unwanted variations in the scene while maintaining the power to tell the difference between object-intrinsic characteristics and scene-accidental conditions. It provides an analysis and references of what are directly observables in a general scene.

1. INTRODUCTION

In image processing, the light as it bounces off the scene is the result of many different causes. In image retrieval and in many other tasks of image processing, we are interested in the light response of only one of them: the light-characteristics of the object embedded in the scene. In this paper, we summarize the causes for the purpose of separating the conditions intrinsic to the object's appearance from the accidental scene.

To handle the problem, one could model the influence of the scene on the appearance of the object, or one could try to capture the intrinsic properties of the object in invariant features. At any rate, modeling scene-specific circumstances has to be bootstrapped by the second approach of invariant characteristics. The invariant approach has the advantage of being less complex at the expense of throwing away essential information. For a complete analysis, neither of the two approaches can be missed. One or another basic invariant observations will bootstrap a model which may invoke more detailed invariant descriptions. In turn, it will bootstrap a model of the scene and so on. In view of this dichotomy, we aim to advance object retrieval in broad domains from tight invariant descriptions.

The aim of invariant descriptions is to identify objects at the loss of the smallest amounts of the information content. If two objects or two appearances of the same object t_i are equivalent under a group of transformations G they are said to belong to the same equivalence class:

$$t_1 \overset{G}{\sim} t_2 \Longleftrightarrow \exists g \in G : t_2 = g \circ t_1 \tag{1}$$

A property f of t is invariant under G if and only if f_t remains the same regardless the unwanted condition expressed by G,

$$t_1 \overset{G}{\sim} t_2 \Longrightarrow f_{t_1} = f_{t_2} \tag{2}$$

In general, a feature with a very wide class of invariance looses the power to discriminate among essential differences. The size of the class of images considered equivalent grows with the dimensionality of G. In the end, the invariance may be so wide that little discrimination among objects is retained. The aim is to select the tightest set of invariants suited for the expected set of non-constant conditions. Hence, in the context of image retrieval, the invariant conditions are to be specified (indirectly) by the user as they reflect the intentions of the user. The oldest work on invariance in computer vision has been done in object recognition. Invariant description in image retrieval is relatively new, but quickly gaining ground. This presentation feeds on the much larger [1, 2, 3].

2. FOLLOW THE LIGHT

In order to analyze what might be observed from a scene in general without too much a priori knowledge about the scene or the objects in the scene, we follow the light. When we ignore the influence of the medium as well as interreflections, the main degrees of freedom are the source, the object, its surroundings and the camera. It is modeled as free parameters in the Schafer-reflection model [4] as follows.

The light starts at the source, where there is freedom to have 1 or more *sources*. Sources may be line sources or even from all directions but these cases can be seen as special cases of multiple point sources. Each source has a *direction* relative to the scene, a spectral composition and intensity. Likewise, the essential free parameters of the camera are the spectral sensitivity, the gain, its *direction* relative to the scene and *distance*. As the spectral content of the source and the spectral sensitivity of the camera have practically the same effect, we take them together under the name *spectral content*. The same holds for the source intensity and the gain of the camera under the name of *intensity*.

For the object, the free parameters can be grouped in

0-7803-6725-1/01/$10.00

free parameters	as seen in scene	directly observables	see
source direction	shadows	one source	[5]
	$\sim$ directions	direction source	
	$\sim$ locations	direction source	
	cast $\sim$	depth order	
source extent	specularity		[6]
spectral content	spectral composition	color source	[7]
intensity	contrast composition	contrast	
camera direction	projection	affine distortion	
camera distance	size composition	depth	[8]
stage setting	occlusion	depth order to view	
	clutter	–	
objects	specularities	color source	
	$\sim$ locations	number sources	
	$\sim$ size	shape source	
	self-shadow	one source	
	shading	direction source	
	$\sim$ maxima	number of sources	[6]

Table 1. Inspired by [9], page 122. What is directly observable from the outer scene? Free parameters of the scene and the single features that can be observed in general without a priori knowledge about the specifics of the scene. Methods listed in the references generally assume sufficiently rich scenes.

the *cover*, ranging from glosse to matte objects. Glosse produces specular reflections. The *albedo* describes the true color of the object, and the *texture* describes the spatial layout of the albedo patterns at its surface. The *touch* of an object describes the 3D-nature of the surface as it introduces a large variability in the perception of the object. In this simplification, the final group of object parameters is grouped under *form*.

For a scene, the one group of parameters left is the stage setting where the objects are placed in the scene in a certain depth order with respect to the light, causing *shadows*, and with respect to the view, causing *occlusion* and *clutter,* preventing the object to be delineated amidst similarly appearing objects.

In listing the main groups of accidental, unknown causes of variation in a general scene, we ignore light-emitting, mirroring, fluid, and transparent objects.

3. THE OUTER SCENE

Given all the sources of variations in a scene, Table, 3 gives an overview of what can be observed about a general scene.

In the reference [5], a method was described capable of discriminating in a natural scene shadow edges from material edges by comparing at each point the invariant description assuming matte and gloss covers. [6] provides a method for the discrimination of ambient light from directed light on the basis of the matte and specular reflection from an object. Similarly, to determine the spectral content of the source (in his case to classify indoor versus outdoor scenes), [7] determines the maximum extent of the spectral content of the image with the spectral content of the two sources. As in a general scene it is likely that there are (unnoticed) specularities somewhere, the spectral components of the source will be seen in the image. Shape from shading not only reconstructs the shape of the object but at the same time reveals the direction of the light source. All these methods indicate that with some effort, something can be said about the spectral content, number and direction of the source from the characteristics of a general scene.

It should be noted that only a few instances of good evidence in the image are needed to decide where the source is and what its spectral contents is. Classification of points as shadow points may fail at many places as long as there is enough evidence pointing at one and the same few sources. Such implies that even if shadow classification relies on material assumptions of Lambertian reflection, the method will be successful in determining the presence of shadow in a general scene as it may be expected that some objects with Lambertian reflection will be approximated by the absence of Fresnel reflection.

Essentially the same approach of surveying the content of the image for composition but then applied to shape is done in [8]. From the diminishing size of affine-invariant descriptors such as fitted ellipses, it is deduced what the depth order in the image is. Again, it may be assumed for a general image that if at the same height in the image the sizes generally are smaller than at the bottom of the image, this will say something about the depth order of the image.

From the table, it is clear that many instances of knowledge about the scene parameters are far from complete. And, we treat the causes as independent factors, ignoring any inter-reflections among them. Especially for closely packed, transparent, mirroring or poly-limbed objects this may not be a valid assumption, but we have to start somewhere.

4. THE INNER SCENE

Table 4 provides a list of free parameters of the object and what can be done to find them.

Photometric invariants [10] and [2] serve as invariant descriptors for the first group of free parameters. In [10], matte patches under white light are described in the c_1, c_2, c_3 color space with $\frac{R-G}{R+G}, \frac{R-B}{B+R}, \frac{G-B}{G+B}$. The color description is invariant for shadow, shading and light intensity and only dependent on the albedo of the object. By computing the ratios $\frac{R_{x_1}G_{x_2}}{R_{x_2}G_{x_1}}, \frac{G_{x_1}B_{x_2}}{G_{x_2}B_{x_1}}, \frac{B_{x_1}R_{x_2}}{B_{x_2}R_{x_1}}$ in the m_1, m_2, m_3 color space the values are invariant under a change in the color of the source while being invariant against shadow and shading. In [2], various color differential invariants are derived computed in the Gaussian color scale space framework, as indicated in table 4. For each feature, the invariance is indicated by a '+'. The number of distinguishable colors is recorded from the 1000 colors of the PANTONE system.

free parameters	as seen on object	constraint on free parameters	directly observables	see
cover albedo				
gloss	specularity	–	cover type H/N	[10]
	$\sim$ locations	–	facing	
	apparent color	–	object color H	[2, 10]
matte	apparent color	white source	object color C	[2, 10]
		color source	constancy N	[2, 10]
texture				
	albedos	–		[11]
	$\sim$ layout	–		[12]
touch				
	meso-highlights	gloss	roughness	
	meso-shadow	one source	roughness	
	meso-shading	matte one source	meso-shape	[13]
form				
matte	macro-shading	one source	shape	
	discontinuity		direction folds	[14]

Table 2. Directly observables in the inner scene. H/N Implies the classification of cover type by combination of results.

	invariant for					
	source				$\sigma = 1$	$\sigma = 3$
	intensity	direction	patch orient	specular		
$\hat{E}$	-	-	-	-	983	1000
$\hat{W}$	+	-	-	-	978	1000
$\hat{C}$	+	+	-	-	820	970
$\hat{N}$	+	+	+	-	757	974
$\hat{H}$	+	+	-	+	461	462

Table 3. From [15]. The trade-off between tightness of the invariance and discriminatory power by showing the number of color patches from 1000 which still can be discriminated. σ denotes spatial scale of the filter. Note that invariance for surface orientation implies invariance for viewing direction and illumination direction.

As expected, the color difference E, with no invariance discriminates all patches, but even for high invariance, color constancy N and generalized hue H keep up considerable discriminatory power.

For the cover the two common types of reflective properties of the object are dull or glosse. For an approximate dull surface reflection is Lambertian its intensity depending on the relative orientation of the patch to the light and the camera. This requires intensity invariant features [10]. From the discussion it is clear that once Lambertian invariance is applied, the result is indistinguishable from intensity variations. Alternatively, an object may have specular reflection, mirroring the spectral properties of the incoming light, [5]. From specular reflections, once we are able to identify them, we can identify the local orientation of the object's surface patch with equal angles towards the camera and the light source, as discussed above under the outer scene. From the diffuse reflection we detect the albedo patterns of the object, provided we have normalized for the light spectra of the source and camera. For the case we do not have a white source, we can say less about the albedo, but we are still capable of retaining some discriminatory power among differently colored objects, [2].

Texture, here loosely defined as the spatial layout of albedo mixtures, in effect is a complex topic yet very helpful for the identification of true object parameters. The mere presence of more colors than just one compensates for the loss of information in the invariant descriptions cited above to uniquely identify an object. Presence of colors in local histograms was studied in [11]. Statistics of local orderings with a large capacity for discriminations were described in [12].

Touch is the group of parameters describing surface roughness, a topic barely touched upon in image retrieval. Small pits in the surface will present itself as a pattern of shadows, whereas smaller variations in the depth will appear as shadings, which may be recovered as such when the other factors such as albedo are assumed constant. Statistics of rough objects illuminated from different directions are given in [13].

The object form will induce large scale shading, which may be recovered by shape from shading approaches. From the table, it is clear that we can have only a limited view on the object and the object properties.

5. AN IMAGE RETRIEVAL SYSTEM BASED ON INVARIANCE

A system for image retrieval on the basis of a visual example should be capable of handling the following accidental conditions of variation in the scene. We discuss geometry-related sources of variation, photometric sources of variations, object related sources and scene related sources of variation:

The unknown location of the object in the image, requires *translation invariance*. This is most commonly encountered. Almost all features in all systems are computed at all locations of the image. An unknown orientation of the object in the plane of the scene requiring *rotation invariant features*. An unknown scale at which the object is observed, requires scale invariant recognition, usually over a range of scales. Scale invariance requires that absolute size can no longer play a role in the recognition. Any references to the area, the number of points (in the histogram) should be made relative to the other areas of the object. The reduction of the 3D-scene to a 2D-view by an unknown projection angle demands the use of affine invariant features. These are most demanding from all factors. Point properties or the (local) size-normalized histogram are most commonly used to design location, orientation and limited scale invariant feature sets.

The photometric sources of variation have been discussed above. Depending on the intent of the user, a source of variation may be the spectral composition of the source. The variations due to the unknown intensity of the illumination will almost implied in other sources of variation unless standardized true color recording is in order as in a museum.

The variability in the object has been described above in extent, it includes cover, albedo, texture, touch and form, all requiring specific sets of invariant features to which table 4 gives a partial solution. Again, it depends on the intent of the user, whether the sources of variation are included in the search. The variation cannot be entirely separated from the photometric variations in the scene.

The influence of the surrounding scene is evident in occlusion. As it wipes out part of the evidence, occlusion prevents the use of whole body shape features. It makes little sense to compute features in an invariant manner while ignoring the difficulties for general object segmentation. Especially the conditions of clutter and occlusion are hard to handle for the average segmentation algorithm. Therefore weak segmentation [1] of identifying points rather than complete segmentations is a sensible way out restricting the similarity to comparing point set groupings [16].

6. CONCLUSION

From the list, we conclude that the use of histograms of local photometric invariant color and texture features tailored to the intent of the user with the query is a powerful and computational efficient approach to image retrieval. Combining shape and color (and in the future texture) all in invariant fashion is a powerful combination as described by [10] where the differential structure of color pairs are stored to identify objects.

When applying invariance for content-based retrieval, the degree of invariance should be tailored to the recording circumstances. Clearly a feature with a very wide class of invariance looses the power to discriminate among essential differences. The aim is to select the tightest set of invariants. What is needed is a complete sets of image properties with well-described variant conditions that they are capable of handling, see [15] and the table 4. This paper was aimed at a further step towards the description of information available from the scene.

7. REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content based image retrieval at the end of the early years," *IEEE Trans. PAMI*, vol. 22, no. 12, pp. 1349–1379, 2000.

[2] J. M. Geusebroek, R. van den Boomgaard, and A. W. M. Smeulders, "Color invariance," *accepted IEEE Trans. PAMI*.

[3] Th. Gevers, "Color based image retrieval," in *Principles of Visual Information Retrieval*, M. Lew, Ed. Springer Verlag, 2001.

[4] S. A. Shafer, "Using color to separate reflection components," *Color Res. Appl.*, vol. 10, no. 4, pp. 210–218, 1985.

[5] Th. Gevers and H. M. G. Stokman, "Classifying color transitions into shadow-geometry, illumination highlight or material edges," in *ICIP*. 2000, vol. 1, pp. 521–525, IEEE Computer Society.

[6] H.J. Andersen and E. Granum, "Classifying the illumination condition from two light sources by color histogram assessment," *JOSA*, vol. 17, no. 4, pp. 667–676, 2000.

[7] G. D. Finlayson, "Color in perspective," *IEEE Trans. PAMI*, vol. 18, no. 10, pp. 1034–1038, 1996.

[8] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Boston, 1994.

[9] M. Baxandall, *Shadows and enlightment*, Yale University Press, New Haven.

[10] Th. Gevers and A. W. M. Smeulders, "Pictoseek: Combining color and shape invariant features for image retrieval," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 102–119, 2000.

[11] B. van Ginneken and J. J. Koenderink, "Texture histograms as a function of irradiation and viewing direction," *Int. J. Comp. Vis.*, vol. 31, pp. 169–184, 1999.

[12] P. Hong, R. Wang, and T. Huang, "Learning patterns from images by combining soft decisions and hard decisions," in *CVPR*. 2000, vol. 1, pp. 78–83, IEEE Press.

[13] J. J. Koenderink, A. J. van Doorn, K. J. Dana, and S. Nayar, "Bidirectional reflection distribution function of thoroughly pitted surfaces," *Int. J. Comp. Vis.*, vol. 31, pp. 129–144, 1999.

[14] C. Schmid and R. Mohr, "Local greyvalue invariants for image retrieval," *IEEE Trans. PAMI*, vol. 19, no. 5, pp. 530–535, 1997.

[15] J. M. Geusebroek, A. W. M. Smeulders, and R. van den Boomgaard, "Measurement of color invariants," in *CVPR*. 2000, vol. 1, pp. 50–57, IEEE Press.

[16] M. Werman and D. Weinshall, "Similarity and affine invariant distances between 2d point sets," *IEEE Trans. PAMI*, vol. 17, no. 8, pp. 810–814, 1995.