A Color Image Classification by Means of Image Transformations

André Smolarz and Philippe Cornu
Laboratoire de Modélisation et Sûreté des Systèmes
Université de Technologie de Troyes
Cedex, France

Abstract

We have previously studied the approaches which permit to associate a texture to an image, and some properties of such transformations. To do so, we use image transformations which act on the pixels of the original image. Some of these transformations have the following property: they statistically mix in a very homogeneous way the pixels of the image. We say that such a transformation is a Quasi-Mixing Transformation (or QM-transformation for short).

For an image, with the use of a QM-transformation we obtain a locally very homogeneous spatial representation of this image (which was previously not homogeneous). This representation appears to be a texture. It is possible to define criteria which permit to quantify the local homogeneity degree of this texture in order, for example, to extract some statistical attributes from a reduced part of the image.

We show that the statistical model of the texture that we obtain allows to define a representation space in order to discriminate the texture classes and thus the colour images.

Introduction

We are interested with the approaches which permit to associate to an image a texture by means of a transformation of the set of pixels of the original image (see Refs. 3 and 4).

Such a transformation is a permutation or a composition of permutations of all the pixels of the given image. So, these transformations are one to one and thus are invertible. For a right choice of the permutation and a right number of iterations, we get a locally very homogeneous spatial representation from an original image which was not previously.

About the last point, if we refer to classical definitions of microscopic textures (see Refs. 5 or 7), we can speak of transformation image $\leftrightarrow$ texture. It is thus possible to define statistical criteria which allow us to quantify the local homogeneity degree of the texture so obtained in order to, for example, extract (compute) some characteristic statistical attributes on a reduced part of the image.

Some of these parts can be used to reproduce the original image with different scale: thus we have a model of the image representation which appear to be a class in a statistical attribute space.

In this paper, in the case of color images, we present some results to illustrate the points mentioned above, and we discuss the perspectives which such an approach may offer in the frame of statistical modelisation and the one of identification or image recognition.

Mixing Systems and Image-Texture Transformations

The definition of the image transformation that we propose here is based on the definition of mixing dynamical systems. These dynamical systems are defined on continuous spaces. Because of lack of place, it is not possible to introduce these systems in this paper. We refer the reader to Ref. 1 for some definitions. We just give an informal but rather intuitive and comprehensive definition (borrowed from Ref. 1) of a mixing transformation: let G be a glass with 90% of Martini and 10% of gin, and let S be a spoon. Then mix the cocktail by turning the spoon in the glass (each turn of the spoon being an iteration). Physically, after a while, one can hope that any part of the cocktail contains 90% of Martini and 10% of gin.

Here, we see the main interest of these transformations: they mix in a very homogeneous way all the elements of the involved space. Always intuitivelly, we can imagine that with a part E of the cocktail, if we apply an inverse transformation restricted to E, we can separate the component in order to have unmixed Martini and gin. It is sure that such a reverse transformation is not possible because " after a while" means when the number of spoon iterations goes to infinity. Nevertheless, we will see later that for digital images (which are finite sets of pixels), this intuitive insight is not erroneous.

Arnold et Avez - see Ref. 1 - give a lot of examples of such mixing transformations, which are defined on the unit square $[0, 1] \times [0, 1]$. In the sequel we will only use one of them, namely the Baker Transform (BT for short). We just mention here that all the examples given by Arnold and Avez are defined on continuous sets. On the other hand, digital images are finite sets of points (pixels). Unfortunatelly, it appears that a transformation of a finite set is never mixing (this fact is intuitivelly comprehensible - see Ref. 2). But for some peculiar mixing transformations like BT, when restricting to finite sets, it remains some mixing like properties: the pixels are statistically well mixed when iterating the transformation.

In order to describe the Baker Transformation, we decompose an iteration in two steps: use first an affine transformation which give an image twice larger and half high (the number of pixels remains unchanged). For Lena (512×512 pixels), this is shown on Figure 1. Thus cut and past the resulting image in order to obtain the image shown on Figure 2 (which has the same size as the original image).

Some Properties of the BT

A digital image is a finite set of points, then the BT is a permutation. Such a transformation is periodic, and the period is a function of the size of the image. For example, for a $2^n \times 2^n$ image the period is 4n. If we stop after 2n iterations we obtain an image which is the original one rotated of 180°. With n iterations the image is perfectly mixed (as we have shown it - see Ref. 3). Figure 3 shows the resulting image after n iterations.

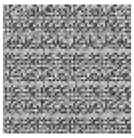
As it can be seen on Figure 3, the transformed image presents a good spatial homogeneity, and thus we can reasonably speak of texture. On the other hand, this kind of transformation works equally for colour images and for grey level images.

Figure 1. Lena during the first iteration of the BT

Figure 2. The image at the end of the first iteration

Original 512×512 image *Transformed image*

Figure 3. From the image to the texture after 9 iterations of the BT



A 64×64 block extracted from the texture (Figure 3)

A 64x64 pixel image computed from the block

Figure 4. From a block extracted from the texture to an image with 18 iterations of the BT

Now, if we divide the $2^n \times 2^n$ texture image in such a way that we get $2^{2(n-m)}$ blocks of size $2^m \times 2^m$ (we have a partition of the texture), with one of these blocks (say the upper left one, but it works for any of the blocks - see Ref. 4) we can compute a $2^m \times 2^m$ image which is semantically identical to the original one (here semantically identical means that we accept a light loss of quality so that the image remains recognizable). This is done with the BT with 3m iterations (or equivalently with m iterations of the inverse of the BT). Figure 4 illustrates this property (here n = 9 and m = 6, so it needs 9 iterations to get the texture and 18 iterations to go from the the extracted block of the texture to the image.

In this context, we can consider that it is possible to use approaches which permit to statistically characterize or give a model of the textures.[6,7,8] The applications we have in mind are characterization, classification and recognition of images. To this end, we have chosen to get a model of the texture by means of computed local statistical attributes which permit a representation of the images as classes. This approach is described in the next section.

Modelisation and Statistical Analysis

Previously, in the case of grey level images (see Ref. 9), we have chosen to get a model of the textures by means of their grey level second order spatial distributions (coocurrences) and to extract from them a set of statistical attributes. Presently, all the images we processed are RGB images and we have computed a coocurrence matrix for each chanel. For the tests that we have carried out and which are presented in the sequel, we have not reduced the number of levels as this is usually recommanded before the coocurrence distribution computation (in the case of grey level images). We have done so because we wanted to preserve the one to one image $\leftrightarrow$ texture relation. For every image, after obtaining the texture with the right number of iterations of the BT (Figure 3), we have computed 3 coocurrence matrices for a distance of 1 pixel in the four main directions, and then, for each matrix, the 12 attributes proposed by Haralick & Al.[7] The coocurrence distributions have been computed on windows of size 32 $\times$ 32 pixels. Each image is described by a set of attribute vectors (64 for a 256 $\times$ 256 image and 256 for a 512 $\times$ 512 image) in a 36-dimensionnal representation space.

We are here in the case of statistical pattern recognition based on the knowledge of a reference or of a set of references (learning set).

Figure 5 shows the 18 images we have used. We have performed a principal component analysis on the whole set of data and we have observed that the main contribution is given by the attribute F7 (sum variances) for the three chanels. Figures 6 and 7-a present the texture classes represented in the 3 planes obtained by combining the three features F7 two by two. Figure 7-b presents the texture classes in the plane (F5-Red , F7-Red), this result shows that we can achieve better discrimination results if we use a set of features larger than the one composed of the three features F7. The recognition results presented in the sequel validate this remark. We can also observe that the classes have a few intraclass and a good interclass dispersion. The small dispersion in each class proceeds from the good texture homogeneity obtained with the TB.

We have performed a recognition test with a priori knowledge. For the learning set, we used 50% of the vectors of each class, and the remainder was used as the test set. In the 3-dimensionnal space composed by the three features F7 and using the Mahalanobis distance, we reached 100% (resp. 99,97%) of good classification results on the learning set (resp. test set). In the 36-dimensionnal space, using the euclidean distance, we reached 100% of good classification results on the learning and test sets. The euclidean distance was used here because the Mahalanobis distance requires the inverse covariance matrix, and it is not possible to compute it in the case of a 36-dimensionnal space with a minimum sample size equal to 32 for the classes corresponding to 256×256 images. The information reduced to a small window produces a good discrimination and we recall that each of these windows corresponds to a representation to a inferior scale of the original image (cf. the properties of the BT and Figure 4).

On the other hand, we can see that the distance proximity between classes do not, in general, agree with a visual or semantic proximity between the corresponding images. This means that we have to choose or define another space representation data to perform recognition and detection of visual similarities between images.

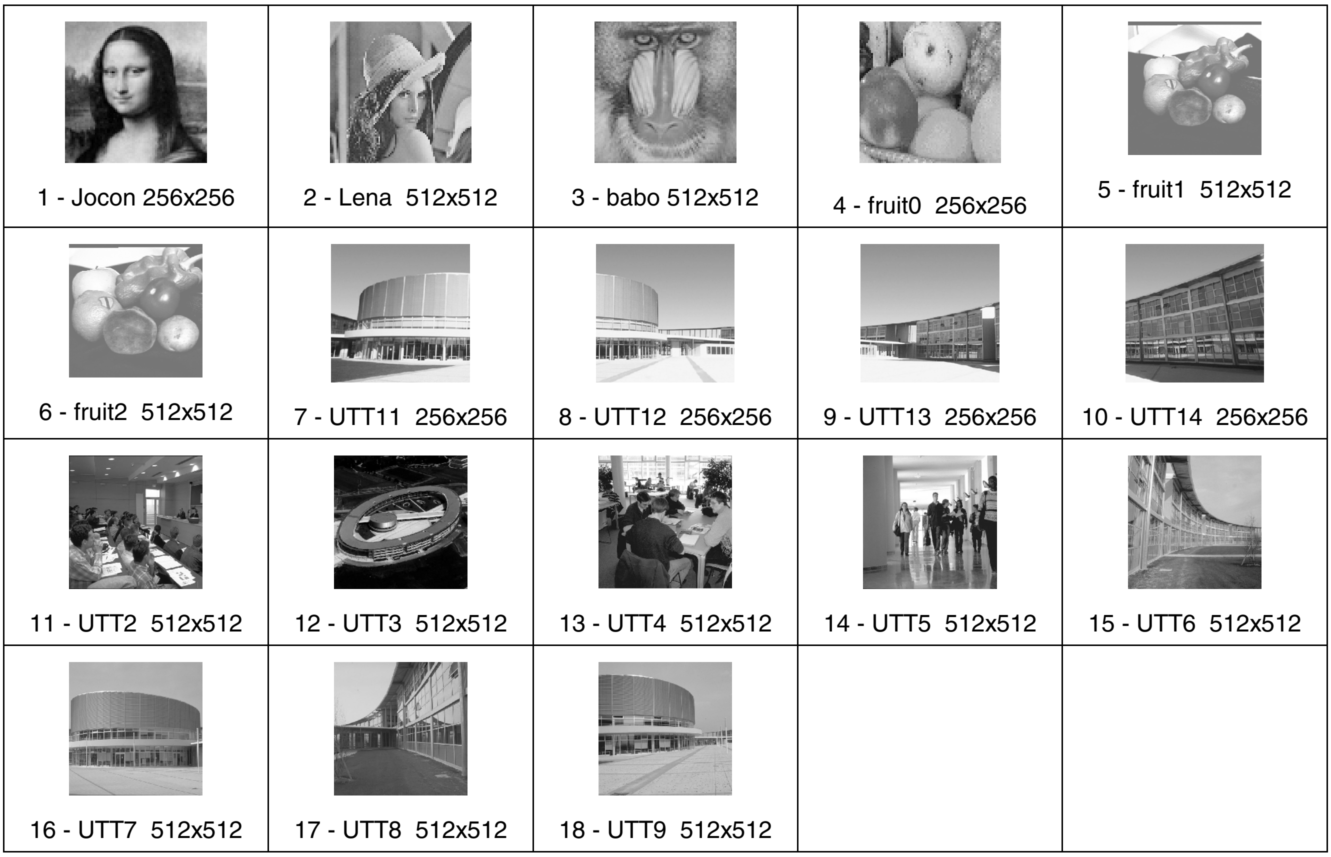


Figure 5. Set of images used for classification and recognition

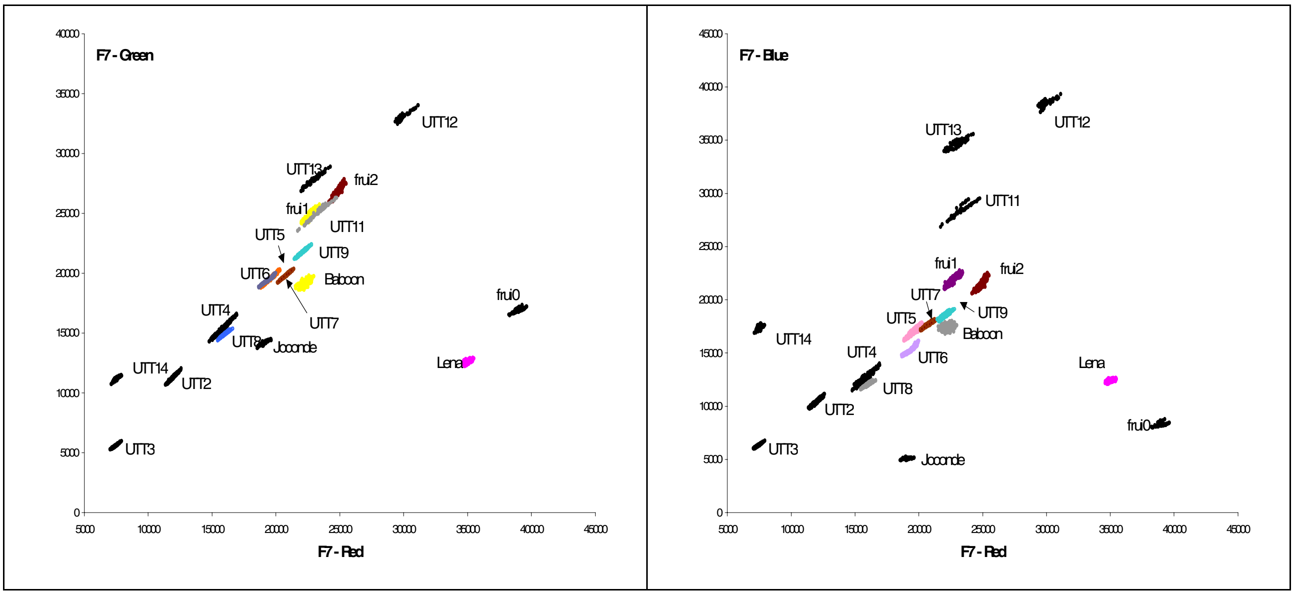


Figure 6. Class image representations - attributes computed on 32x32 windows of textures (see Figure 3)

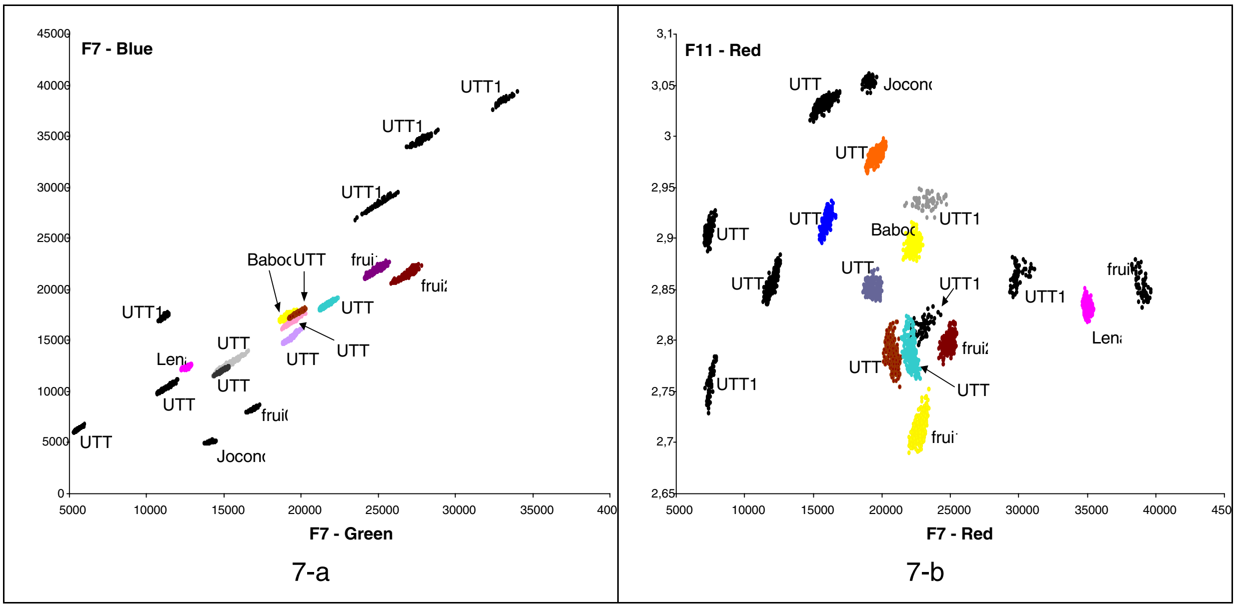


Figure 7. Class image representations - attributes computed on 32x32 windows of textures (see Figure 3)

Conclusion

We have introduced a transformation (BT) which allows to associate in a one to one way a texture to a color or to a grey level image. We have shown that the statistical model of the textures so obtained permits to define a representation space in order to discriminate texture classes and thus the images. On the other hand, we have also previously observed (see Ref. 9) that if the original images were (grey level) textures, the BT gives rise to a class representation rather more accurate, the spatial interclass relations being preserved. From a statistical pattern recognition point of view, this latter point shows that the BT permits to improve the discrimination between classes. The case of color textures has not yet been addressed.

Beside the fact that the BT has some rather interesting properties, we actually search some other quasi-mixing transformations (with better mixing properties) which may hopefully produce better results.

On the other hand, we have seen that proximity relations between classes generally do not correspond to semantic proximity between images. If we wish by example to use such an approach to classify images with resemblance criteria, we have to define some textures attibutes which allow a good discrimination and which are near from the visual content of the images. We also actually work on these aspects in the case of color images and of grey level images.

References

1. V. I. Arnold, A. Avez, "Problème ergodiques de la mécanique classique", Monographies Internationales de Mathématiques Modernes n°9, Gauthier-Villars, Paris, 1967.
2. P. Billingsley, "Ergodic Theory and Information", John Wiley & Sons, Inc., New-York, 1965.
3. Ph. Cornu, A. Smolarz, Caractérisation de la signature texturelle d'une image, actes du 18ème colloque GRETSI'01, 10-13 septembre 2001, Toulouse, Tome 1, p. 359-362.
4. Ph. Cornu, A. Smolarz, "Caractérisation d'images par textures associées", Journée Thématique 2000 du GDR-PRC ISIS : Signal et image au service de la sécurité dans la société de l'information, CNRS - Meudon, 10 Octobre 2000.
5. A. Gagalowicz, "Vers un modèle de textures", Thèse d'état, université Paris VI, mai 1983.
6. M.M. Galloway, "Texture Analysis using Gray Level Run Lengths", Computer Graphics & Image Processing 4, pp 172-179, 1975.
7. R. M. Haralick, K. Shanmugam, Its'hak Dinstein, "Textural Features for Image Classification", IEEE Trans. on systems, man & cybernetics, Vol. SMC-3, N° 6, pp 610-621 Nov 1973.
8. A. Smolarz "Discrimination de textures à l'aide de caractéristiques statistiques locales entre blocs de pixels". XXXèmes journées de Statistique 25-29 mai 1998 Rennes. pp 518-521.
9. A. Smolarz, Ph. Cornu, Transformation d'images pour la modélisation et l'extraction de caracteristiques statistiques, Actes de la journée thématique "Coopération Analyse d'Images et Modélisation" (CoopANIM) EEA-ISIS, 14 juin 2001, Lyon, pp. 18—23.

Biography

André Smolarz received his Thèse de Docteur-Ingénieur from the Université de Technologie de Compiègne (UTC-France) in 1982. Since 1994 he is Associate Professor at the Université de Technologie de Troyes (UTT - France) where he teaches probabilities, statistics and pattern recognition. His work in the Laboratoire de Modélisation et Sûreté des Systèmes (LM2S) focused on texture and image analysis.

Philippe Cornu received his Thèse de troisième cycle in Mathematics from the Université de Montpellier (France) in 1981 and his Thèse d'Etat in Computer Science from the Université de Technologie de Compiègne (UTC - France) in 1987. Since 1997 he is Associate Professor at the Université de Technologie de Troyes (UTT - France) where he teaches programming languages and computer security. His work in the Laboratoire de Modélisation et Sûreté des Systèmes (LM2S) focused on image analysis and on image processing from a discrete point of view.