Chapter 4

Recognition based on multiresolution decomposition

In this Chapter we address the problem of color texture classification and present results on practical problems. The central idea is to combine color and texture information through the multiresolution decomposition of each channel in order to take as feature vector the energies and cross correlations of the coefficient images. However, this simple approach can be materialized in many different ways, as several decisions have to be taken, each one allowing multiple choices: the multiresolution decomposition scheme (e.g. Mallat's, *à trous*, wavelet packets), the subspaces base family (and within it, which specific base), number of decomposition levels, space for color representation and finally, the classification features to be computed from the decomposition. Instead of simply trying some possibilities and take the best one, we have assessed a very large number of combinations, trying to find out which are the important and the non-relevant issues with regard to the classifier performance. In addition, we propose three image models as a framework for color texture classification, depending on how texture is combined with color. This allows us not only to initially select the appropriate types of features but also to reduce the number of classification parameters so that the training set does not need to be large. This framework has been successfully applied to two specific machine vision problems, namely, the sorting of ceramic tiles into perceptually homogeneous classes and the recognition of metalized paints for car refinishing. Also, it has been applied to the classification problem of petrographical marble images solved in the previous Chapter with a different approach.

4.1 Introduction

In this work we present a study on the wavelet decomposition and classification of color textures. It was prompted to solve an industrial machine vision problem, namely, the on-line sorting of polished ceramic tiles. Later, we broach another application, paint identification from microscopy images for car refinishing. We will see that the solution to the first application also fits very well the second one. Finally, we use the

same approach to attempt the classification of marble images used in the previous Chapter.

From the point of view of computer vision, we are addressing in all cases a problem of color texture representation and classification. The objective is to devise a numerical representation of images that captures both the color and texture features. In the present case, we are interested in the benefits of this representation for classification purposes, but it may also be useful in other contexts like color texture synthesis and database image indexing.

The whole application process consists of the following steps, being the slanted items those directly associated with the method of color texture analysis we propose in this work:

- image acquisition
- *change of color space representation*
- *multiresolution decomposition*
- *feature extraction*
- supervised classification

We will focus on the justification and performance of different choices for color representation spaces and multiresolution decomposition schemes. Our aim is to assess all possible combinations in terms of minimum classification error over a relatively large set of samples. Furthermore, we want to provide a sound explanation in terms of *why* each choice achieves its result. This is done in the context of models that we propose to analyse color texture.

This Chapter is organized as follows. Section 4.2 reviews previous work on joint computational representations of color and texture visual cues, including wavelet transforms of multichannel images. Next section describes the planned experiments: the choices for color space, wavelet transform scheme and the classification features derived from them, which we will combine and assess. In Section 4.4 we introduce the problem of ceramic tile classification and the main results obtained. Section 4.5 briefly deals with the same issues for the second case study of paint recognition. Section 4.6 is a link between the classification based on structural parameters treated in Chapter 3 and the proposed methodology presented here. Finally, Section 4.8 describes the conclusions and discusses future work.

4.2 Related work

Color texture representation is a current topic in computer vision. Although both are properties of a surface, these two visual cues have been usually studied separately. One reason is that while color is a point feature given by the value of a pixel in several bands or channels, texture has been modeled as a spatial relationship of the point with its neighbours within each channel. An excellent review of approaches used in computer vision to deal with the texture representation problem can be found in [79], whereas and introduction to color representation is given in [103].

The study of color texture representations has received increasing attention in the last years. The objective of many researchers is to find co-joint representations of spatial and chromatic information which capture the spatial dependence (in particular, correlation) *within and among* spectral bands [14, 33, 24]. One of the most frequent approaches is the construction of a feature vector mixing gray level texture features and color features [33]. Another one is to extend classical texture models, such as Markov Random fields and the autocorrelation function, in order to deal with multichannel images [24, 37]. Other works, like [31], convert RGB values into a single code from which texture measurements are computed as if it were a gray scale image. Spatio-chromatic representations are computed in [14, 29] over the smoothed Laplacian of the image. Other works have been influenced by known perceptual mechanisms of the human visual system like Gabor filters [97, 46].

In parallel, multiresolution texture analysis has come to age thanks to the setting of a sound theoretical basis for wavelet transforms and filter banks. Recent works on texture incorporate color as and additional image dimension [97, 62, 94]. This has been applied to analysis but also to synthesis [22, 39] and texture classification [97, 46].

A color texture analysis based on a multiresolution decomposition representation normally involves to make up two decisions: the selection of the decomposition scheme to perform the texture analysis and the definition of a space to represent color. A general framework for image decomposition is to apply a bank of filters. Gabor filter banks and wavelet transforms are two common approaches found in the literature.

The simplest way to extend them to cope with color images is to filter or transform each channel (RGB for instance) independently. However, some authors propose to represent color in other spaces such as the opponent color space [46], inspired in biological evidences of the human visual system. Both works start from similar color representations, followed by different texture analysis methods. We are going to devote some attention to them, as they are closely related to our study.

The first one [97] uses the orthogonal wavelet decomposition and calculates the energy, e_i^k, of each detail level and the cross terms between different channels at the same detail level, c_i^{kl}:

$$e_i^k = \int \left(d_i^k(u)\right)^2 du \tag{4.1}$$

$$c_i^{kl} = \int d_i^k(u) d_i^l(u) du \ , \tag{4.2}$$

where u denotes spatial coordinates, i the decomposition level, k and l are channel indexes. Thus, d_i^k is the detail at level i of the channel k. In this specific case, d is an image of detail coefficients of a orthogonal wavelet decomposition, but it can be seen also as one of the outputs of a filter bank.

The second work [46] uses a set of Gabor filters where the response at different levels and channels is analysed. A biological model is implicit in this scheme due to the use of Gabor filters and to the extraction of the information between channels following the opponent color model. Energies at each level of every channel (terms e_i^k of Eq. (4.1) for all i and k) are calculated, but also the energies associated to the

inhibition between channels at different levels

$$I_{ij}^{kl} = \int \left(d_i^k(u) - d_j^l(u)\right)^2 du \ , \tag{4.3}$$

where d_i^k are now the responses of a Gabor filter bank. If we expand the inhibition terms of Eq. (4.3) we obtain the energies e_i^k, e_j^l and a cross term that could be expressed as $-2c_{ij}^{kl}$, using the notation of Eq. (4.2). Therefore, both papers are using a similar representation.

To end this review, we want to mention a sound comparative study on the performance of texture classification algorithms by Randen and Husøy [78]. Like us, they want to assess combinations of wavelet decompositions and features, including additional filter bank schemes. However they test them only on gray level images. But the main shortcoming of their study with regard to ours is that they work with clearly distinct textures, that is, a subset of the Brodatz, Meastex and Vistex collections. Conversly, we are trying to differentiate among textures much more visually similar, as they come from the same industrial process (at least in the tiles case), this being a much tougher problem, as real problems usually are.

There are a few previous works to be considered in the specific subject of tile inspection. Some research effort has been devoted to the detection of other kinds of defects like cracks and spots. Only in [11, 70] the same problem of tile color texture classification is addressed. The authors try to solve it taking as features statistical measurements on the color histogram. Therefore, results are poor in the event of overall similar color but different textural aspect, as it happens in our samples. Better results were obtained by performing a color segmentation prior to an analysis of blob features [6].

4.3 Multiresolution color texture classification

4.3.1 Color spaces

It is a common practice to use color representation that try to decorrelate information across channels, thus reducing the number of meaningful classification features. However, we will consider other choices, such as conversion from color to intensity and no transformation at all, in order to compare them with the decorrelation transforms. Therefore, the envisaged color spaces/transforms are:

C.a Color to gray level conversion by simple averaging of the R, G and B channels. Hence, only intensity is taken into account. This would make sense in images where texture is the only relevant feature for classification.

C.b Raw RGB values. That is, no transformation is applied to the image provided by the camera and frame grabber. In many applications this is sufficient to introduce the color information and classify successfully.

C.c Ohta color space. This space is a good approximation of the results of Karhunen-Loève transform over a big set of natural images [42]. This is obtained

through the base that best decorrelates the spectral information of a large set of color images. It is similar to the transformation used in [97], where they use also a generic K-L transform. This color space transformation is given by the following fixed linear transform:

$$\begin{bmatrix} 0.3 & 0.3 & 0.3 \\ 0.5 & 0.0 & -0.5 \\ -0.25 & 0.5 & -0.25 \end{bmatrix} \begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix} . \tag{4.4}$$

It does not decorrelate spatially but somehow gets three new channels weighing each one by its real contribution when describing the input data with the new base.

C.d Specific Karhunen-Loève transform. Now, the base which achieves the maximum spectral decorrelation is sought, but over the specific training set of the application. In our case, it is the set of images for each class and model. It is similar to the previous case but adjusting the projection axis to the data.

To fix ideas and for the sake of simplicity, we will illustrate concepts of this section with figures of the tile problem. Figure 4.1 shows the former four transforms for a 128×128 region of a tile.

There are a big amount of possible color spaces transformations [69] that can be applied in this problem at this preliminary stage that were not evaluated. Instead of an exhaustive search among the different color spaces we reduce it to a few selected cases. Firstly, in the **C.a** case we tend towards a reduction of data to gray level (this can be see as a reduction to the first eigenvector in a K-L transform). In the second case, **C.b**, we do not manipulate data at all as most of the color texture analysis in the literature do. Next, we try to decorrelate the color information by means of a K-L transform: first, in case **C.c**, by a global transformation for a big number of images; and finallly, in case **C.d**, a K-L transform adapted to the data.

Among all the possible spaces not used here we emphasize the opponent color space because it is physiologically motivated. Proposed by Hering in the 19th century and updated by Hurvich and Jameson in the 1960's this theory assumes three sets of receptor systems, red-green, blue-yellow and black-white, bearing in mind the process observed in the human visual system. Althought this specific color space transformation has not been used here the opponent theory is implied in this and the following processes (decomposition and feature extraction) because the reponses of different channels and the cross-information between channels are evaluated and it is related to the opponent features Eq. (4.3).

4.3.2 Decomposition scheme and bases

The decomposition scheme is application dependent. Thus, for time critical applications, an orthogonal scheme as the proposed by Mallat with a reduced number of levels is generally preferred. Conversely, in images with high frequency content in the middle zone of the spectrum, a wavelet packet scheme should be better a priori because it allows to focus the analysis on the levels where the important information

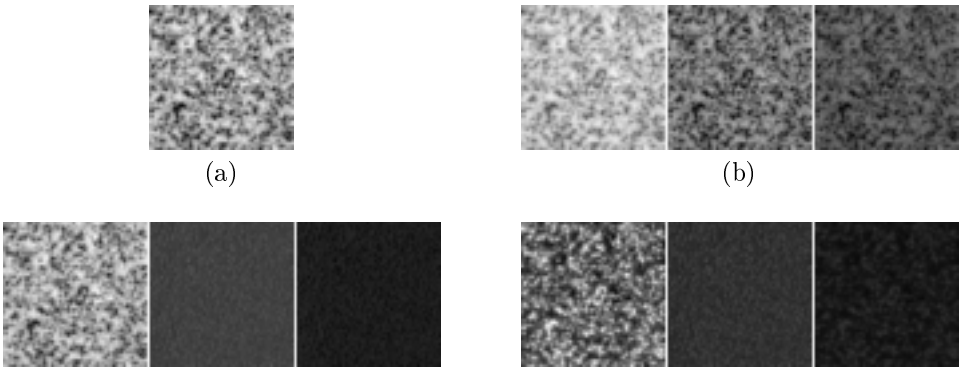

Figure 4.1: Color spaces: (a) gray level, (b) RGB, (c) general Karhunen-Lòeve transform (Ohta color space), and (d) specific Karhunen-Lòeve transform. Images have been linearly contrast enhanced for the sake of visualization.

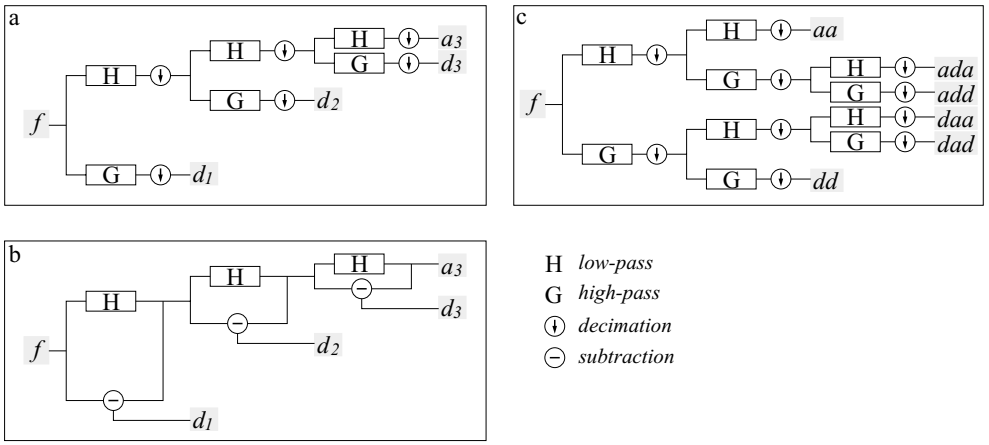

Figure 4.2: Decomposition schemes of a 1D signal f into detail (d) and approximation (a) coefficients for: a) Mallat's, b) *à trous* and c) wavelet packet transforms.

is. Likewise, in images which exhibit a regular behavior and without a privileged direction, an isotropic and symmetric decomposition makes more sense, but then it must be non-orthogonal and redundant like the *à trous* decomposition. Therefore, the following wavelet transforms have been considered:

D.a Multiresolution analysis with Mallat's algorithm [62].

D.b *À trous* algorithm [89]. Opposite to **D.a** and **D.c**, it is a non-orthogonal and hence redundant transform.

D.c Wavelet packets transform [106] using a few fixed tree structure patterns.

As in the previous color space selection, there are some other multiresolution decomposition schemes that have not been evaluated. Maybe, among them, Gabor is the most outstanding and widely studied in the literature. We have reduced the decomposition schemes to the three mentioned wavelet cases.

In addition to the wavelet transform scheme, a suitable base must be selected. There are many families of bases, each having different properties like symmetry, orthogonality and regularity (related to the number of vanishing moments). This adds still a new dimension to the search space in which we want to minimize the classification error. We have studied widely the first problem, classification of tiles, to limit the number of proofs to do, in subsequent experiments, in order select the best parameters, bases, decomposition schemes. Several resolutions, bases and schemes has been evaluated as we see in Appendix A (Table A.2). In order to cut down the number of tests, we have fixed the base family for each scheme after a number of trials. Accordingly, Mallat's multiresolution analysis and wavelet packets transform are performed with Daubechies orthogonal bases and the *à trous* decomposition uses B-spline bases. Figure 4.2 summarizes the decomposition scheme followed by each transform in a 1D setup, the 2D extension of these algorithms has been described in Section 2.2.7. Figure 4.3 shows an example of these three decompositions over the R channel of a tile. As we can see in the figure the *à trous* decomposition is a redundant transformation; each new detail level increase the total amount of data in a quantity equal to the initial image. We can see how wavelet and wavelet packed based on ortogonal transformations have not this behavior; the amount of data does not increase.

4.3.3 Feature extraction

Once the decomposition has been performed, we need to compute a vector of features. In the literature of wavelet texture analysis two types of features are mostly used: energy and entropy. They are applied to the coefficients of the approximation and details at each level, though in some works cross energies (correlation signatures in our terminology) of details at different levels are also computed. Joint entropy [96, 19] of couples of details or approximations at different levels and/or channels could also be computed and assessed.

In our applications both features had a similar performance. Actually, energy attained less than 1% improvement on the classification error over entropy, at least when features were restricted to be the energies of details and approximation for each channel (terms e_i^k, see bellow in Eq. (4.5)). Although this slight improvement is not a sufficient reason to dicard entropy results, we have restricted our study to energy related features in order to reduce the number of possible features.

The terms we will compute for the analysis stage are the energy and the cross correlation between levels and channels. We call all of them *correlation signatures* like in [97]:

$$c_{ij}^{kl} = \int d_i^k(u) d_j^l(u) du \ . \tag{4.5}$$

Note that c_{ij}^{kl} also include the energy terms because $e_i^k = c_{ii}^{kk}$.

The number of features provided by the former three decompositions grows rapidly

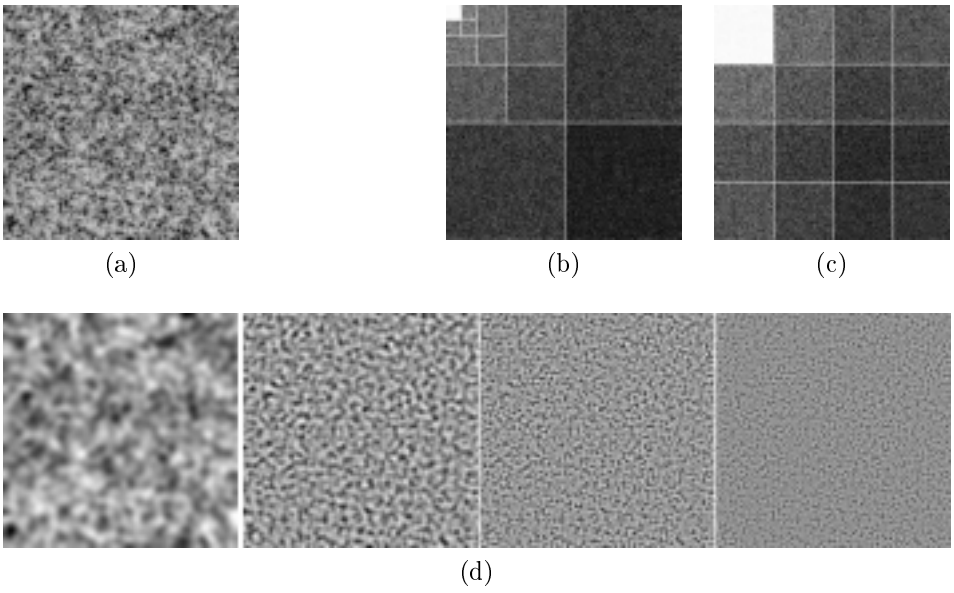

Figure 4.3: Decomposition examples: (a) 256×256 image region of the red channel of a tile. (b) A four level Mallat's wavelet transform. (c) One of the 13 possible two level wavelet packet decompositions (leafs of the tree at the second level). A logarithm transformation has been applied on images (b) and (c) for the sake of visualization. (d) Approximation and detail levels of the *à trous* decomposition. All of them have been contrast maximized separately.

as the number of levels increases. For instance, a three-levels Mallat's wavelet transform of a RGB image gives rise to 30 images (1 approximation plus 29 detail images) on which 306 correlation signatures of images of the same size are possible. In a well devised supervised classifier, when the number of discriminant features increases, the performance is enhanced. However, if this number is too large with regard to the size of the training set, the classifier just learns to succeed over this set but it is not able to generalize. Therefore, we should keep small the length of the feature vector, namely, the number of signature terms. For this reason, we propose to test the following choices, illustrated in Fig. 4.4:

F.a Compute only the energy terms: $c_{ii}^{kk}\ \forall i, \forall k$. This is the most frequent choice in the literature.

F.b Calculate all correlation signatures between levels but only within the same channel: $c_{ij}^{kk}\ \forall i, j, \forall k$.

F.c Calculate all correlation signatures between channels but only within the same level: $c_{ii}^{kl}\ \forall i, \forall k, l$. This is the approach taken in [97].

F.d Collect all possible correlation signatures between channels and levels: $c_{ij}^{kl}\ \forall\, i, j, \forall\, k, l$. In order to select relevant features, we take into account the former

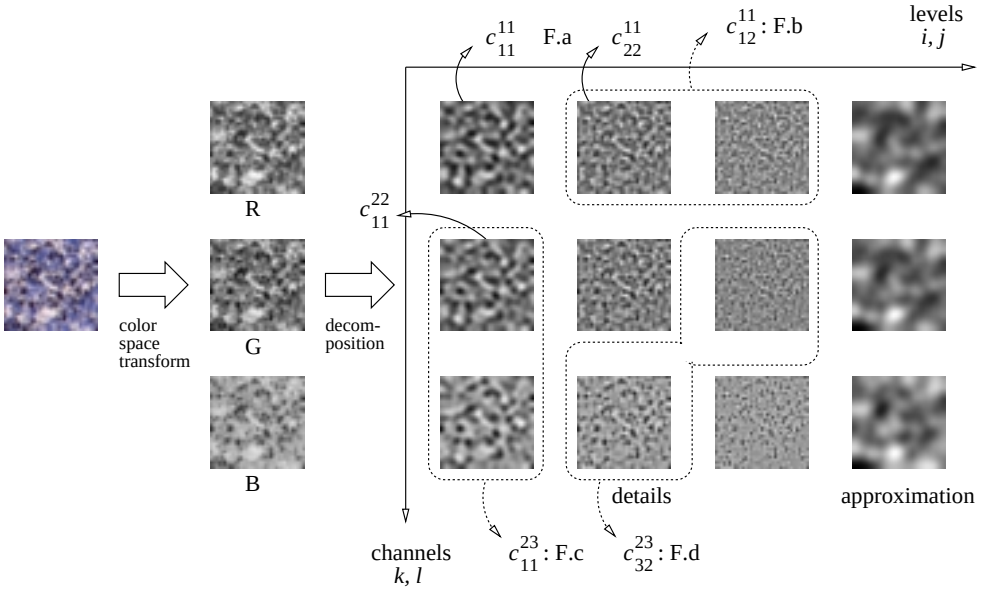

Figure 4.4: Features are selected among four types of correlation signatures.

observation of Section 4.2 which related some correlation signatures with the inhibition energies of the opponent color model.

Mallat's and wavelet packets transforms in which the decomposition tree has different levels can not be combined with options **F.b** and **F.d** unless we perform a specific transformation. The reason is that coefficient images at different detail levels have different size due to decimation, thus, it is not possible to cross-correlate them. If were necessary this problem can be solved removing the decimation step or scaling the coeficient image in a deep level to the appropiate size of the level to be compared.

4.3.4 Models

As we stated before, our approach to color texture classification is to first select a suitable space for color representation, a multiresolution decomposition scheme of the image represented in this space, and finally a set of discriminant features derived from this decomposition. However, it does not make sense to try every possible combination of choices for the three former items. Instead, we must select them according to and **image model** which explains how texture is related or mixed with color. We propose the following three models:

M.a Images resulting from the addition of a gray level texture plus a uniform background color. Thus, only energy terms **F.a** from approximation and detail coefficients at different levels and **F.b** make sense. Furthermore, as this model in fact assumes a same texture for each channel, the former features must be

computed just over one of the channels or the intensity image (mean of R, G and B).

M.b Now, we assume that each channel contributes with a different texture to the final visual aspect of the image. But we further suppose that these textures are statically independent. Therefore, only **F.a** and **F.b**, this time over each channel, are candidates to be discriminant features with regard to a classification task. This is the model used in [39] for texture synthesis.

M.c Conversely to M.b, we suppose now that textures along each channel are dependent, and in particular linearly dependent. Thus, besides **F.a** and **F.b**, correlation signatures between approximation or detail coefficients of different levels and channels, **F.c** and **F.d**, must be taken into account as potentially discriminant features.

4.3.5 Classification method

The classification method is a nonparametric discriminant analysis. In order to classify new samples we need a set of prototypes representing each possible class. Afterwards, the distance between the sample and each class can be calculated and the most similar class assigned. Given that classes are not known a priori, we need some method to learn the prototypes from a set of samples.

One of the methods that fits our constraints is that of Fisher discriminant functions, because, without any a priori knowledge of data, it is able to select the best representation maximizing the ratio between the inter-class covariance and the intra-class covariance [65]. A linear transform W, is applied to the feature vector $\boldsymbol{x}$ of a particular image obtaining a new representation, $\boldsymbol{y} = W^t\boldsymbol{x}$, in a space where the discriminant capacity is maximized.

The linear transformation, W, which optimizes the discrimination is obtained by calculating the most significant eigen vectors of the matrix $S_w^{-1}S_b$, assuring that the following ratio is maximized:

$$\frac{W^t S_b W}{W^t S_w W} \ ,$$

where S_w is the within data sparse matrix, defined as:

$$S_w = \sum_{i=1}^{c} \sum_{\boldsymbol{x}_k \in c_i} (\boldsymbol{x}_k - \boldsymbol{\mu}_i)(\boldsymbol{x}_k - \boldsymbol{\mu}_i)^t \ ,$$

where c is the number of possible classes and c_i is the set of vectors that are used as learning samples in the i class. The S_b matrix is the between class sparse matrix, which is defined as:

$$S_b = \sum_{i=1}^{c} N_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^t \ ,$$

where $\boldsymbol{\mu}_i$ is the mean vector of the samples of the i class, N_i is the number of learning samples in the i class and $\boldsymbol{\mu}$ is the global mean vector.

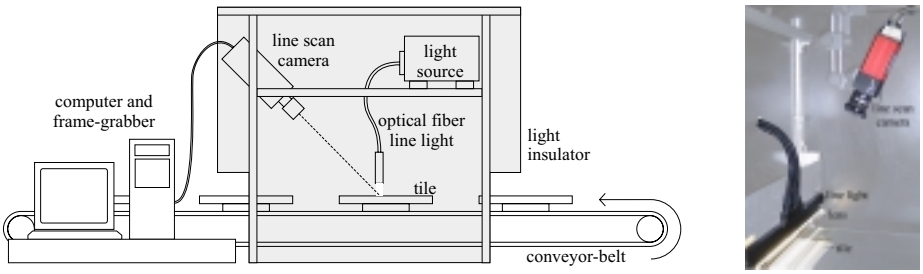

Figure 4.5: Setup of the tile inspection system and detail of the camera and illumination system.

From an image of a tile, we extract its feature vector, $\boldsymbol{x}$, and we assign it to class j if

$$|W^t\ \mathbf{x} - W^t\boldsymbol{\mu}_j| < |W^t\boldsymbol{x} - W^t\boldsymbol{\mu}_i| \qquad \forall\, i \neq j\ .$$

Further details on the classifier can be found in [28, 65].

4.4 Sorting of ceramic tiles

4.4.1 The problem

Tile manufacturing needs of pigments and clay which are mixed, melted, sprayed on to the tile substrate, and finally baked. Unavoidable variations in the pigments color, temperature, humidity and pressure conditions provoke subtle visual variations of the tile aspect when tiles are placed on the floor, one next to other. These visual changes are due to small differences in color and texture, and are seen as defects by customers. A system was thus needed to automatically sort tiles from a given model into perceptually homogeneous classes. At present, several trained workers at the end of the production line perform this task. In each production line only a model of tiles is produced. Thus, classification must be done among classes of each model and not among models. As it is a tedious, time-consuming and subjective task, an automated system is needed.

We have built a system prototype to acquire and analyze images from tiles (see scheme of Fig. 4.5). Tile images are acquired with a 3 CCD digital line scan camera which yields 10 bits per channel. This allows us to distinguish color details invisible to the human eye, though a very stable lighting is required. We have designed a line light system which integrates several halogen sources and optical fiber light guides. In addition, we adapt the spectral content of the light to the camera CCD sensitivity by placing a set of color filters in front of the lens. Tiles move on a conveyor-belt with controlled speed under the linear camera, and this allows us to adjust the vertical resolution of the images to be the same as in the horizontal direction. The horizontal resolution is 5 pixels/mm, and it is given by the camera height above the conveyor belt and the lens fixed focal length.