Video Segmentation through Multiscale Texture Analysis

Miguel Alemán-Flores and Luis Álvarez-León

Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran Canaria
Campus de Tafira, 35017, Spain
{maleman,lalvarez}@dis.ulpgc.es

Abstract. Segmenting a video sequence into different coherent scenes requires analyzing those aspects which allow finding the changes where a transition is to be found. Textures are an important feature when we try to identify or classify elements in a scene and, therefore, can be very helpful to find those frames where there is a transition. Furthermore, analyzing the textures in a given environment at different scales provides more information than considering the features which can be extracted from a single one. A standard multiscale texture analysis would require an adjustment of the scales in the comparison of the textures. However, when analyzing video sequences, this process can be simplified by assuming that the frames have been acquired at the same resolution. In this paper, we present a multiscale approach for segmenting video scenes by comparing the textures which are present in their frames.

1 Introduction

In this paper, we present a method for video segmentation based on the distribution of the orientation of the edges. We use the results of the multiscale texture analysis described in [1] and study the behavior of natural textures in order to find the transitions between the different video scenes. To this end, we estimate the gradient in every point of the region and build an orientation histogram to describe it. This allows performing satisfactory classifications in most cases, but some of them are not properly classified. A multiscale analysis of the textures improves the results, considering the evolution of the textures along the scale. In natural textures, the changes produced when a certain scene is observed at different distances introduce new elements which must be taken into account when comparing the views. This texture comparison technique is applied to video segmentation by considering those intervals within which the energy is low enough to be considered as normally evolved video sequences.

The paper is structured as follows: Section 2 shows how textures can be described and classified through their orientation histograms. In section 3, multiscale analysis is introduced to improve the classification method and some considerations are analyzed in natural textures. Section 4 describes the application of multiscale texture comparison to video segmentation. Finally, in section 5, we give an account of our main conclusions.

2 Texture Description and Classification

In order to describe a texture in terms of the edges which are present in it, we must estimate the magnitude and the orientation of the gradient in every point of the region. With these values, we can build an orientation histogram which reflects what the relative importance of every orientation is. We first calculate an initial estimation for every point using the following mask for the horizontal component x_i and its transpose for the vertical component y_i:

$$\frac{1}{4h}\begin{pmatrix} -(2-\sqrt{2}) & 0 & (2-\sqrt{2}) \\ -2(\sqrt{2}-1) & 0 & 2(\sqrt{2}-1) \\ -(2-\sqrt{2}) & 0 & (2-\sqrt{2}) \end{pmatrix} . \tag{1}$$

Using the structure tensor method, the orientation of the gradient at a certain point can be estimated by means of the eigenvector associated to the lowest eigenvalue of the matrix in (2), whereas the magnitude can be approximated by the square root of its highest eigenvalue. We first convolve the image with a Gaussian to increase the robustness of the approximations. By adding the magnitude in the points with the same orientation, we can build an orientation histogram for each texture. These histograms are normalized, so that the global weight is the same for all of them.

$$\begin{pmatrix} \sum_{i=0}^{N} y_i^2 & -\sum_{i=0}^{N} x_i y_i \\ -\sum_{i=0}^{N} x_i y_i & \sum_{i=0}^{N} x_i^2 \end{pmatrix} . \tag{2}$$

In order to compare two textures, an energy function is built, in which the Fourier coefficients of both histograms are analyzed. A change in the orientation of a texture will only cause a cyclical shift in the histogram. For this reason, the Fourier coefficients are modified as follows: let f_n and g_n be the orientation histograms of length L corresponding to the same texture but shifted a positions, i.e. the texture has been rotated an angle $\theta = 2\pi a/L$, and let f_k and g_k be the k^{th} Fourier coefficients of these histograms, then $f_k = g_k e^{-i\frac{2\pi k a}{L}}$.

In addition, the fact that the number of discrete orientations used for the histograms is constant as well as the normalization of the weights make the lengths of the signals and the total weight equal in both textures. Due to the fact that the higher frequencies are more sensitive to noise than the lower ones, a monotonic decreasing weighting function $w(.)$ can be used to emphasize the discrimination, thus obtaining the following expression, in which the first terms have a more important contribution than the last ones:

$$E(a) = \sum_{k=1}^{\frac{L}{2}} w\left(\frac{2k}{L}\right)\left(f_k - g_k e^{-i\frac{2\pi k a}{L}}\right)\left(f_k - g_k e^{-i\frac{2\pi k a}{L}}\right)^* . \tag{3}$$

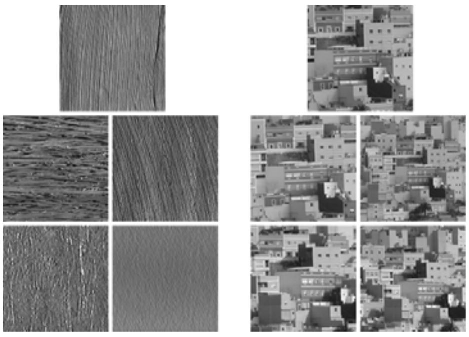

Fig. 1. Results of searching for similar textures for a texture in database 1 and a texture in database 2

To test this technique, we have used two sets of textures contained in two databases. The first database has been made publicly available for research purposes by Columbia and Utrecht Universities [2] and consists of different materials. The second one corresponds to different natural scenes acquired at several distances. In Fig. 1, we show some results of the application of the technique explained above. From the image databases, one is selected and the five images which produce the lowest energies are shown.

The orientation histograms extracted from the textures describe how the different orientations are quantitatively distributed across the region which is studied, but they provide no information about the spatial neighborhood of the pixels with a certain orientation. Thus, a completely noisy image, in which all orientations are found in approximately the same proportion, but in a disordered way, would generate a similar histogram as a circle, where the orientation increases gradually along its outline. This forces us to search for a certain technique which complements the information provided by this kind of histograms in order to enhance their recognition capability.

3 Multiscale Texture Analysis

The interpretation of the information we perceive from the environment depends on the scale we use to process it. The multiscale analysis approach has been successfully used in the literature for texture enhancement and segmentation (see [3] and [4] for more details).

A multiscale analysis can be determined by a set of transformations $\{T_t\}_{t\geq 0}$, where t represents the scale. Let I be an image, i.e. $I : \Omega \longrightarrow \Re$, where Ω is the domain where the image is defined. We will consider that $\Omega = \Re^n$, $I \in H^2(\Omega)$ (I and ∇I have finite L^2 norm) and $I_t = T_t(I)$ is a new image which corresponds to I at a scale t. For a given image I, to which the multiscale analysis is applied, we can extract a histogram $\{h_i^t\}_{i=0,..,L-1}$ which determines the distribution of the orientations of I at scale t. In this case, the normalization of the values within a histogram is performed with respect to the initial addition. In order to compare the histograms of two images, the scale must be first adjusted.

3.1 Gaussian Multiscale Analysis

We will use a Gaussian filter, whose properties are described in [5] and [6]. In one dimension, this process can be quantized as follows, where the scale t is related to the standard deviation σ according to the expression $2t = \sigma^2$:

$$(x * K_t)_m = \sum_{n=-\infty}^{\infty} x_n \frac{1}{\sqrt{4\pi t}} e^{-\frac{(m-n)^2}{4t}} \quad . \tag{4}$$

Given a signal f, the result of convolving f with the Gaussian filter K_t is equivalent to the solution of the heat equation, given by $\partial u/\partial t = \partial^2 u/\partial x^2$, where $u(t,x)$ is the solution of the equation, using f as the initial data $(u(t,s) = K_t * f(x))$. Considering this relationship, a discrete version of the heat equation can be used to accelerate the approximation of the Gaussian filtering (see [7] for more details), which results in a recursive scheme in three steps for each direction.

This process will be performed by rows and by columns in order to obtain a discrete expression for a two-dimensional Gaussian filtering. Making use of the features of the Gaussian kernels, the result of applying a Gaussian filter with an initial scale t can be used to obtain a Gaussian filtering of the initial image for a different scale with no need to start again from the input.

3.2 Multiscale Orientation Histogram Comparison

We must take into account that, for a certain texture, the use of different resolutions forces us to apply Gaussian functions with different standard deviations, thus requiring an adaptation stage. To do that, we extract the evolution of the magnitude of the gradients at different scales and we use them to compare the textures. Even if the quantitative distribution of the orientations may be alike for different textures, the spatial distribution will cause a divergence in the evolution, so that the factors will differ.

One of the properties of the Gaussian filtering is the relationship between the resolution of two images and the effects of this kind of filters. In fact, the result of applying a Gaussian filter with standard deviation σ to an image with resolution factor x is equivalent to applying a Gaussian filter with standard deviation $k\sigma$ to the same image acquired with a resolution factor kx.

Given two textures, I_0 and I'_0, we will estimate the scale factor k using the normalized evolution of the addition of the norm of the gradient, that is, we will use:

$$\phi(I_0, \Omega, t) = \frac{\sqrt{\int_\Omega |\nabla I_t|^2}}{\sqrt{\int_\Omega |\nabla I_0|^2}} \quad . \tag{5}$$

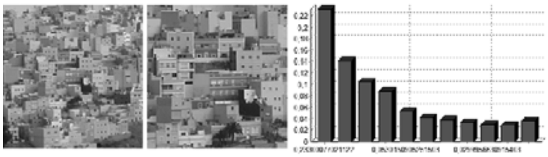

Fig. 2. Comparison of two similar textures at different scales

It is well known (see for instance [5]) that $\phi(I_0, \Omega, t)$ is a decreasing function with respect to t and $Lim_{t\to\infty}\phi(I_0, \Omega, t) = 0$. On the other hand, if $I'_0(x, y) = I_0(kx, ky)\ \forall(x, y) \in \Omega$, then $\phi(I_0, \Omega, t) = \phi(I'_0, k\Omega, k^2t) = \phi(I'_0, \Omega, k^2t)$, considering that the texture is periodically repeated.

Consequently, in order to estimate a scale factor k between two textures I_0 and I'_0, we will compare the functions $\phi(I_0, \Omega, t)$ and $\phi(I'_0, \Omega, t)$. Let $r^1_n = \phi(I_0, \Omega, (\sigma_n)^2/2)$ and $r^2_n = \phi(I'_0, \Omega, (\sigma_n)^2/2)$ be the ratios obtained for two textures at scale $\sigma_n = n\sigma_0$, the best adjusting coefficient k to fit the series of r^2_n to that of r^1_n, both consisting of N terms, can be obtained as follows: We first fit a value $0 < h < 1$ and we interpolate the values in the series r^1_n and r^2_n to obtain two new series σ^1_n and σ^2_n which estimate the scales for which the ratios $(1, 1-h, 1-2h, 1-3h, ..., 1-(N-1)h)$ are obtained. In other words, we estimate the scale where $\phi(I, \Omega, \left(\sigma^1_n\right)^2/2) = 1 - nh$. We must point out that, if $nh < 1$, then σ^1_n and σ^2_n are well-defined, because $\phi(I, \Omega, t)$ is a decreasing function with respect to t and $Lim_{t\to\infty}\phi(I_0, \Omega, t) = 0$. With these values, we minimize the following error to obtain the scale factor k:

$$e(k) = \frac{1}{N}\sum_{i=0}^{N-1}\left(\sigma^1_i - k\sigma^2_i\right)^2 \quad . \tag{6}$$

We can study how the energy obtained when comparing the orientation histograms evolves as we apply a Gaussian filtering to the textures. We use the adjusting factor k to relate the scales to be compared and we obtain the energies for the comparison of the histograms at N different scales.

Figure 2 shows the results of comparing two images corresponding to similar textures, acquired at different distances. As observed, not only the initial energy is low, but also the subsequent energies, obtained when comparing the images at the corresponding scales, decrease when we increase the scale. On the other hand, Fig. 3 shows the comparison of two images of different textures. The energies, far from decreasing, increase from the initial value.

3.3 Resolution Adjustment in Natural Scenes

We have extracted the evolution of the square of the gradient across the image for all the textures in the second database, in which different natural scenes

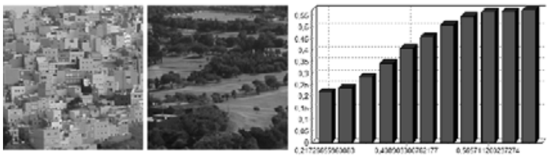

Fig. 3. Comparison of two different textures

have been acquired at different distances. With these values, we have calculated a ratio for every couple of pictures in the database. Instead of observing a great variability in the ratios according to the different natures and distances, they are very close to 1 in most cases. The fact that certain particular elements appear when we approach them, while other global elements disappear, thus generating new gradients while other ones are eliminated, makes the total addition similar, and the information, in terms of changes existing in the image, is approximately constant. In fact, the mean ratio for the comparison of two textures, considering in each case the ratio which is lower than 1, is 0.91975, with standard deviation 0.06190. In artificial textures, a change in the resolution produces a change in the evolution of the addition of the squares of the gradients and no additional information is added, thus generating more variable ratios.

4 Video Segmentation

The multiscale comparison of natural textures described above has been used to segment video sequences by finding the transitions in which the texture histogram undergoes a great change. On the assumption that, when a scene finishes and a new one starts, the textures in the frames are quite different, the energies obtained when comparing them will be significant and the transition can be located.

If we force the system to be sensitive enough to avoid overlooking any scene transition, the threshold which determines from which value a change is considered as significant may be too low to avoid including some intra-scene changes as transitions, thus reducing the specificity. At the same time, the transitions can be either abrupt, i.e. a scene finishes in frame n and the new scene starts in frame $n + 1$, or soft, i.e. there is a diffusion, shift, or any other effect to go from a scene to the following. The latter type forces us to compare frames which are not consecutive in order to detect the change. But this might include more intra-scene changes as transitions. Thus, a multiple temporal interval is needed.

We have used a set of videos and reports provided by researchers from the Universidad Autónoma de Madrid [8]. Human observers have signaled the frames where a transition is found, and we have compared these values with the frames where the energy is higher that a certain threshold. We have used four versions of every frame: the original image and the image after the application of a Gaussian

Table 1. True transitions (TT) and false transitions (FT) located using original scale analysis for time interval 10 (OSA10), multiscale analysis for time interval 10 (MSA10) and multiscale analysis for combined time intervals 10 and 5 (MSA10-5). Number of frames: 2500, number of transitions: 21, number of comparisons: 249

Method	TT Detected	FT Detected	% of FT
OSA10	21	40	18
MSA10	21	27	12
MSA10-5	21	23	10

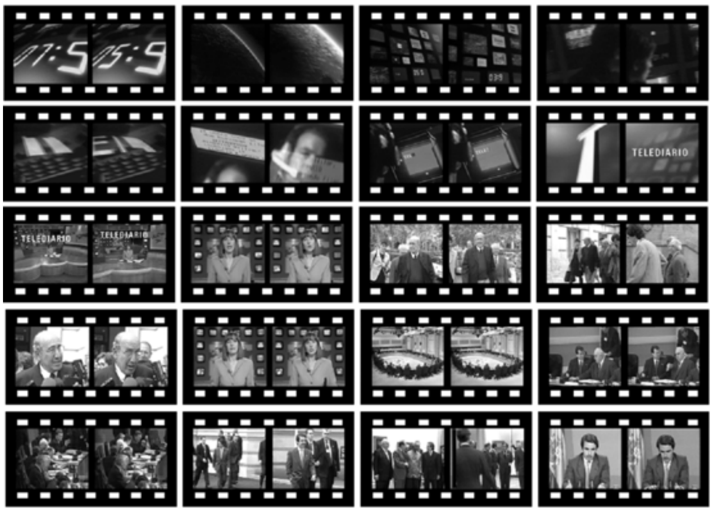

Fig. 4. Example of scenes and transitions detected in a video sequence. Every couple of images corresponds to the initial and final frames of a scene

filter with $\sigma = 1, 5$ and 10. The best results have been obtained using the mean of the two intermediate values for $\sigma = 0, 1, 5$ and 10.

If we use an interval of 10 frames in texture comparison in order to determine where a transition occurs, we are able to detect all actual transitions in the sequence of video frames. However, 18% of normal changes, i.e. those which occur between frames of the same scene, are labelled as transitions, since there is a considerable evolution of the elements in them. If we consider a combination of the energies for σ = 0,1,5 and 10, these false transitions are reduced to 12%. Furthermore, if we select the candidates to be transitions for a temporal interval of 10 frames and we analyze them with a temporal interval of 5 frames, we can refuse some of them considering the changes as normal intra-scene evolutions and the false transitions are reduced to 10%. Table 1 shows a comparison of the results using these methods. Figure 4 shows the initial and final frames of different scenes extracted for a video sequence.

5 Conclusion

In this paper, we have presented a new approach to video sequence segmentation based on a multiscale classification of natural textures. By using the structure tensor, we have obtained an estimation of the gradient in every point of the textures. The extraction of orientation histograms to describe the distribution of the orientations across a textured region and the multiscale analysis of the histograms have produced quite satisfactory results, since the visual similarity or difference between two textures is much more reliably detected by the evolution of the energies resulting when comparing the histograms at different scales.

We have observed how the ratio for the adjustment of the scales is not far from 1 when natural images are considered, since the information contained in them changes qualitatively, but not as much quantitatively. The need for a high sensibility, in terms of transitions detected in order to avoid overlooking them, produces a decrease in the specificity, in such a way that certain false transitions appear as such when the energy is extracted. However, the comparison at different scales and using different temporal intervals reduces significantly these misconstrued normal changes while preserving the right ones.

The promising results obtained in the tests which have been implemented confirm the usefulness of the multiple comparison of the images, since they endow us with a much more robust discrimination criterion.

References

1. Alemán-Flores, M., Álvarez-León, L.: Texture Classification through Multiscale Orientation Histogram Analysis. Lecture Notes in Computer Science, Springer Verlag **2695** (2003) 479-493
2. Columbia University and Utrecht University. Columbia-Utrecht Reflectance and Texture Database. http://www.cs.columbia.edu/CAVE/curet/.index.html
3. Paragios, N., Deriche, R.: Geodesic Active Regions and Level Set Methods for Supervised Texture Segmentation. International Journal of Computer Vision **46:3** (2002) 223
4. Weickert, J.: Multiscale texture enhancement, V. Hlavac, R. Sara (Eds.), Computer Analysis of Images and Patterns, Lecture Notes in Computer Science Springer Berlin **970** (1995) 230-237
5. Evans, L.: Partial Differential Equations. American Mathematical Society (1998)
6. Lindeberg, T.: Scale Space Theory in Computer Vision. Kluwer Academic Publishers (1994)
7. Álvarez, L., Mazorra, L.: Signal and Image Restoration Using Shock Filters and Anisotropic Diffusion. SIAM J. on Numerical Analysis **31:2** (1994) 590-605
8. Bescós, J.: Shot Transitions Ground Truth for the MPEG7 Content Set. Technical Report 2003/06. Universidad Autónoma de Madrid (2003)