Chapter 7

Image retrieval and analysis

Erkki Oja, Jorma Laaksonen, Jukka Iivarinen, Markus Koskela, Ramūnas Girdziušas, Jussi Pakkanen, Ville Viitaniemi, Zhirong Yang, Rami Rautkorpi, Mats Sjöberg, Hannes Muurinen

7.1 Content-based image retrieval by self-organizing maps

Erkki Oja, Jorma Laaksonen, Markus Koskela, Ville Viitaniemi, Zhirong Yang, Mats Sjöberg, Hannes Muurinen

Content-based image retrieval (CBIR) has been a subject of intensive research effort for more than a decade now. It differs from many of its neighboring research disciplines in computer vision due to one notable fact: human subjectivity cannot totally be isolated from the use and evaluation of CBIR systems. In addition, two more points make CBIR systems special. Opposed to such computer vision applications as production quality control systems, operational CBIR systems would be very intimately connected to the people using them. Also, effective CBIR systems call for means of interchanging information concerning images' content between local and remote databases, a characteristic very seldom present, e.g., in industrial computer vision.

PicSOM

The methodological novelty of our neural-network-based CBIR system, PicSOM [1, 2], is to use several Self-Organizing Maps in parallel for retrieving relevant images from a database. These parallel SOMs have been trained with separate data sets obtained from the image data with different feature extraction techniques. The different SOMs and their underlying feature extraction schemes impose different similarity functions on the images. In the PicSOM approach, the system is able to discover those of the parallel SOMs that provide the most valuable information for each individual query instance.

Instead of the standard SOM version, PicSOM uses a special form of the algorithm, the Tree Structured Self-Organizing Map (TS-SOM) [3]. The hierarchical TS-SOM structure is useful for large SOMs in the training phase. In the standard SOM, each model vector has to be compared with the input vector in finding the best-matching unit (BMU). With the TS-SOM one follows the hierarchical structure which reduces the complexity of the search to $O(\log n)$. After training each TS-SOM hierarchical level, that level is fixed and each neural unit on it is given a visual label from the database image nearest to it.

Self-organizing relevance feedback

When we assume that similar images are located near each other on the SOM surfaces, we are motivated to exchange the user-provided relevance information between the SOM units. This is implemented in PicSOM by low-pass filtering the map surfaces. All relevant images are first given equal positive weight inversely proportional to the number of relevant images. Likewise, nonrelevant images receive negative weights that are inversely proportional to their number. The relevance values are then summed in the BMUs of the images and the resulting sparse value fields are low-pass filtered.

Figure 7.1 illustrates how the positive and negative relevance responses, displayed with red and blue map units, respectively, are first mapped on a SOM surface and how the responses are expanded in the low-pass filtering. As shown on the right side of the figure, the relative distances of SOM model vectors can also be taken into account when performing the filtering operation [4]. If the relative distance of two SOM units is small, they can be regarded as belonging to the same cluster and, therefore, the relevance response should easily spread between the neighboring map units. Cluster borders, on the other hand, are characterized by large distances and the spreading of responses should be less intensive.

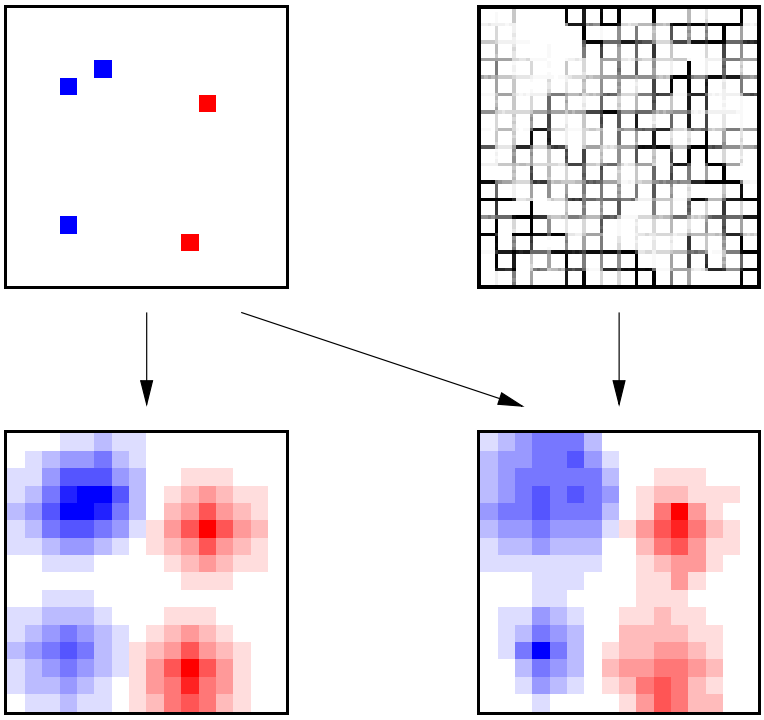

Figure 7.1: An example of how positive and negative map units, shown with red and blue marks on the top-left figure, are low-pass filtered. Two alternative methods exist; either we ignore (bottom-left figure) or take into account (bottom-right figure) the relative distances between neighboring SOM model vectors. In the top-right figure, the relative distances are illustrated with gray level bars so that a darker shade of gray corresponds to a longer relative distance between two neighboring map units.

Finally, the set of images forming the result of the query round is obtained by summing the relevance responses or *qualification values* from all the used SOMs. As a result, the different content descriptors do not need to be explicitly weighted as the system automatically weights their opinions regarding the images' similarity and relevance.

User interaction feature

Relevance feedback can be seen as a form of supervised learning to adjust subsequent query rounds by using information gathered from the user's feedback. It is essential that the learning takes place during one query, and the results are erased when starting a new one. This is because the target of the search usually changes from one query to the next, and so the previous relevances have no significance any more. This is therefore *intra-query* learning.

Relevance feedback provides information which can also be used in an *inter-query* or *long-term* learning scheme. The relevance evaluations provided by the user during a query session partition the set of seen images into relevant and nonrelevant classes with respect to that particular query target. The fact that two images belong to the same class is a cue for similarities in their semantic content. This information can be utilized by considering the previous user interaction as metadata associated with the images and use it to construct a *user interaction* or *relevance feature*, to be used alongside with the visual features. This method was presented and experimented with in [5]. An example of a resulting SOM is illustrated in Figure 7.2. In the figure, a 16×16-sized SOM trained with user interaction

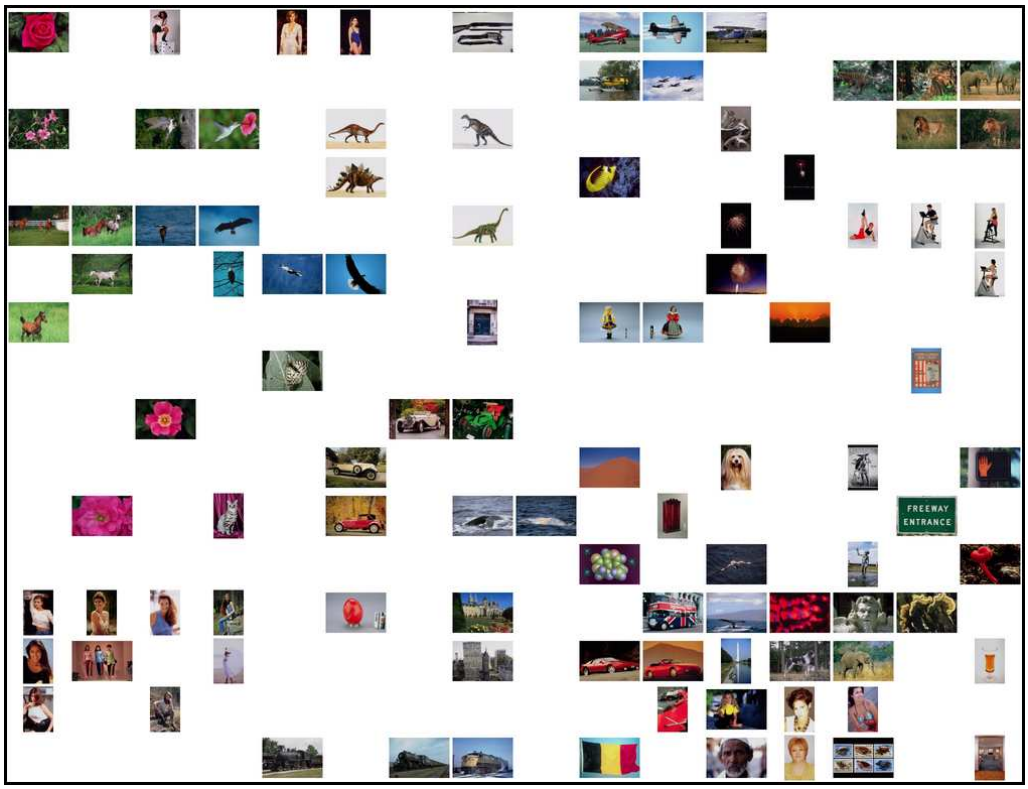

Figure 7.2: The image labels of a 16×16-sized SOM trained with user interaction data.

data is shown. It can be observed that images with similar semantic content have been mapped near each other on the map.

In some cases, the image database may also contain manually assigned or implicit annotations. These annotations describe high-level semantic content of the image and often contain invaluable information for retrieval purposes. Therefore, it is useful to note that the user-provided relevance evaluations discussed above are notably similar to these annotations. In particular, keyword annotations can be seen as high-quality user assessments and the presented method can be readily utilized also for these annotations.

Use of segmented images

The general problem of image understanding is intrinsically linked to the problem of image segmentation. That is, if one understands an image one can also tell what the different parts of it are. Segmentation thus seems to be a natural part of image understanding. As the objects and entities in the images usually hierarchically decompose into sub-objects, the segmentations defined by the images form hierarchical segmentation trees.

For an automatic system segmenting an image is never trivial and the results seldom correspond to the real objects in the picture. But even so segmentation may be useful in CBIR, because different, visually homogeneous regions somehow characterize the objects and scenes in the image. The existing approaches differ mainly in the fashion the segment-wise similarities are combined to form image-wise similarities used in the retrieval. The hierarchical structure of segmentations is usually ignored.

The PicSOM system uses the results of unsupervised segmentation by generalizing the original algorithm so that not only the entire images but also the image segments are seen

as objects in their own right. The segments are also considered to be sub-objects of the images they are a part of. The relevance feedback process is modified so that when an image is marked as relevant all its sub-objects (segments) are also marked as relevant. Then, after calculating qualification values for all the objects on the different TS-SOMs, the qualification values of all the sub-objects are summed to their parent objects. Finally, the values obtained from different maps are again summed up to form the final image-wise qualification values. In the algorithm, different levels of segmentation hierarchies can be considered simultaneously. The approach can be used also for salient image parts that have been identified by detecting interest points in the images.

In practical applications and preliminary experiments we have observed that for most of the used ground truth image classes, the retrieval precision obtained by using both entire and segmented images together excels over that obtained by using either ones alone. In forthcoming systematic experiments we will further explore the usefulness of these various approaches for exploiting the spatial decomposition of the images into subparts on one hand, and combining the approaches on the other hand.

In addition to interactive CBIR, the PicSOM system has been used off-line to investigate the connection between semantic concepts and results of unsupervised segmentation. In particular, we have considered the keyword focusing problem where the system autonomously learns to attribute annotating keywords to specific segments. The system discovers the segment-keyword connections from example images where the annoting keywords are given on per-image basis. An example of a relevance focusing setting is shown in Figure 7.3 where pieces of clothing are localised in a database of fashion images.

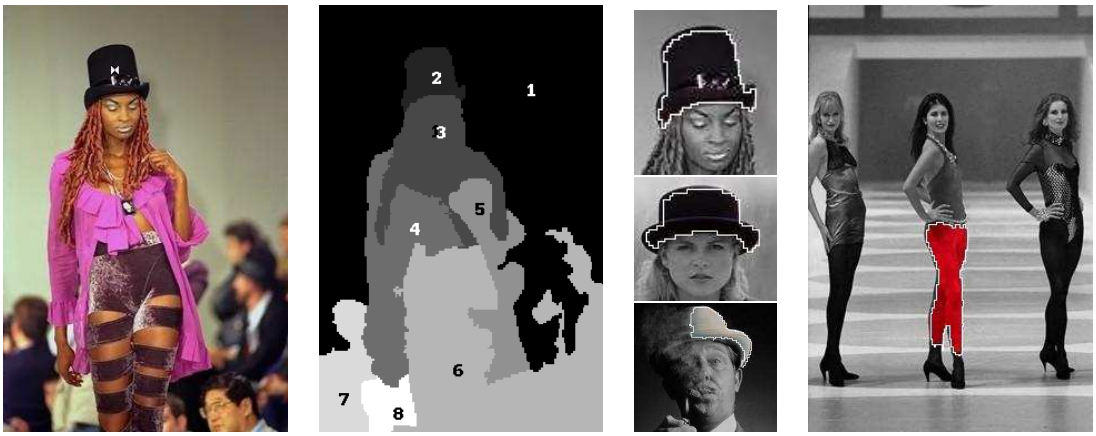

Figure 7.3: Focusing keywords “hat” and “red trousers”.

Interactive facial image retrieval

Most existing face recognition systems require the user to provide a starting image, which is however not practical e.g. when searching for a criminal based on a witness’ power of recall. Interactive facial image retrieval, which are mainly based on learning the relevance feedback from the user, is a potential approach to address this problem.

The early appearance of the first subject hit is critical for the success of the retrieval. Unlike content-based image retrieval systems based on general images, the query precision on facial images suffers from the problem of extremely small sizes of the *subject classes* [6]. If only images that depict the correct person are regarded as relevant, many pages of only non-relevant images would be displayed. Because the negative responses from the user in the early rounds provide little semantic information, the iteration progresses in a

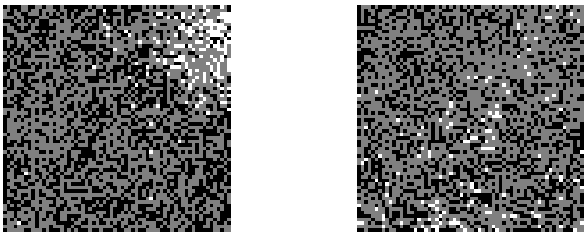

Figure 7.4: A 64×64 DSOM example shown on the left. The white points represent the class members of *mustache_yes* and black points for *mustache_no*. The normal SOM without discriminative preprocessing is shown on the right for comparison.

nearly random manner.

We have proposed a novel method which adaptively learns partial relevance during the interactive retrieval. First, we extended the PicSOM CBIR system by replacing the membership of a subject class with that of a semantic class as the new relevance criterion until the first subject hit appears. Second, we applied supervised learning as a preprocessing step before training the Self-Organizing Maps so that the resulting SOMs have stronger discriminative power. Figure 7.4 visualizes an example *Discriminative Self-Organizing Map* (DSOM) of two semantic classes, *mustache_yes* and *mustache_no*. The empirical results on FERET database [7] have shown that the number of displayed images can be significantly reduced by employing these two strategies [8].

Different ways for obtaining a DSOM exist. We applied Fisher's linear discriminant analysis (LDA) as a preprocessing step before training the SOMs. More advanced methods for extracting the discriminative components of data are now under development.

Multimodal hierarchical objects

The basic ideas of CBIR can be expanded to the more general concept of content-based *information* retrieval if we also consider objects of other types than images, for example text, video, and multi-media objects in general. The only restriction is that we must be able to extract some kind of low-level statistical feature vectors from the new data objects. For example, text data similarity can be evaluated with the n-gram method.

A possibly even more important development has been the incorporation of the context and relationships of data items as hierarchical objects in PicSOM. For example a web page with text, embedded images and links to other web pages can be modeled as a hierarchical object tree with a web page object as parent and the text, links and images as children.

The PicSOM system has been extended to support such general multimodal hierarchical objects and a method for *relevance sharing* between these objects has been implemented [9]. This means that the relevance assessments originally received from user feedback will be transferred from the object to its parents, children and siblings. This means for example, if we want to search for an image of a cat from a multimedia message database, we can let the system compare not only the images but also the related textual objects. If the reference message text contains the word "cat" we can find images which are not necessarily visually similar, but have related texts containing the same keyword.

The hierarchical object paradigm has been applied to many problems, such as webpage structures and e-mail message retrieval. Most recently the PicSOM group took part in the NIST TRECVID 2005 workshop where we successfully applied these methods to video retrieval [10]. In addition to the hierarchical object method we also applied *class models* as a way of representing semantic concepts. The class models were previously used for image group annotation in [11], whereas in the TRECVID context we are interested

in video shots of the test collection that have the highest likelihood of being relevant to a given concept.

The multi-part hierarchy for video shots used for indexing the TRECVID 2005 collection is illustrated in Fig. 7.5. The video shot itself is considered as the main or parent object. The keyframes (one or more) associated with the shot, the audio track, and ASR/MT text are linked as children of the parent object. All object modalities may have one or more SOMs or other feature indices, and thus all objects in the hierarchy may have links to a set of associated feature indices.

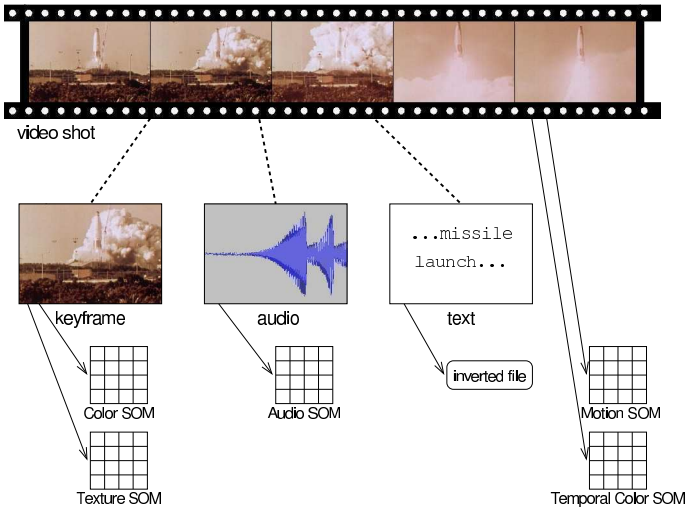

Figure 7.5: The hierarchy of video and multimodal SOMs.

Class models were created by utilizing semantic classifications of the training set videos provided by NIST, such as a list of clips showing an explosion or fire. Feature vectors of objects belonging to such class models can be mapped as impulses onto to the corresponding SOM surfaces, providing an estimate of the true class distribution. In retrieval the sign of the impulses were adjusted to represent relevant (positive) and non-relevant (negative) concepts. The sparse value fields on the maps are low-pass filtered to spread the information which increases the probability that objects mapped to nearby SOM units will be retrieved.

In the high-level feature extraction task we applied the class model representations to find video clips representing different concepts, applying an SFS-type feature selection scheme for each concept separately. The search tasks were run with the PicSOM system by using predefined search topics in three ways: with user interaction, with some manual tweaking of the parameters and without user interaction. To our delight, the PicSOM results compared very well with the other systems taking part in the TRECVID 2005 evaluation. For example, in the automatic search tasks, the PicSOM results were clearly above the median in all but one topic.

References

[1] J. T. Laaksonen, J. M. Koskela, S. P. Laakso, and E. Oja. PicSOM - Content-based image retrieval with self-organizing maps. *Pattern Recognition Letters*, 21(13-14):1199–1207, November 2000.

[2] J. Laaksonen, M. Koskela, and E. Oja. PicSOM – Self-Organizing Image Retrieval with MPEG-7 Content Descriptors. *IEEE Transactions on Neural Networks*, 13(4): 841-853, July 2002.

[3] P. Koikkalainen and E. Oja. Self-organizing hierarchical feature maps. In *Proc. International Joint Conference on Neural Networks*, vol. II, pages 279-285, Piscataway, NJ, 1990.

[4] M. Koskela, J. Laaksonen, and E. Oja. Implementing Relevance Feedback as Convolutions of Local Neighborhoods on Self-Organizing Maps. In *Proc. International Conference on Artificial Neural Networks*, pages 981-986. Madrid, Spain. August 2002.

[5] M. Koskela and J. Laaksonen. Using Long-Term Learning to Improve Efficiency of Content-Based Image Retrieval. In *Proc. Third International Workshop on Pattern Recognition in Information Systems*, pages 72-79, Angers, France, April 2003.

[6] Zhirong Yang and Jorma Laaksonen. Interactive retrieval in facial image database using Self-Organizing Maps. In *Proceedings of IAPR Conference on Machine Vision Applications (MVA 2005)*, pages 112–115, Tsukuba Science City, Japan, May 2005.

[7] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22:1090–1104, October 2000.

[8] Zhirong Yang and Jorma Laaksonen. Partial relevance in interactive facial image retrieval. In *Proceedings of 3rd International Conference on Advances in Pattern Recognition (ICAPR 2005)*, pages 216–225, Bath, UK, August 2005.

[9] Mats Sjöberg and Jorma Laaksonen. Content-based retrieval of web pages and other hierarchical objects with Self-Organizing Maps. In *Proceedings of 15th International Conference on Artificial Neural Networks (ICANN 2005)*, pages 841–846, Warsaw, Poland, September 2005.

[10] Markus Koskela, Jorma Laaksonen, Mats Sjöberg, and Hannes Muurinen. PicSOM experiments in TRECVID 2005. In *Proceedings of the TRECVID 2005 Workshop*, pages 262–270, Gaithersburg, MD, USA, November 2005.

[11] Markus Koskela and Jorma Laaksonen. Semantic annotation of image groups with Self-Organizing Maps. In *Proceedings of 4th International Conference on Image and Video Retrieval (CIVR 2005)*, pages 518–527, Singapore, July 2005.

7.2 Content-based retrieval of defect images

Jussi Pakkanen, Jukka Iivarinen, Rami Rautkorpi

A need for efficient and fast methods for content-based image retrieval (CBIR) has increased rapidly during the last decade. The amount of image data that has to be stored, managed, browsed, searched, and retrieved grows continuously on many fields of industry and research.

In this project we have taken a noncommercial CBIR system called PicSOM, and applied it to several databases of surface defect images. PicSOM has been developed in our laboratory at Helsinki University of Technology to be a generic CBIR system for large, unannotated databases. We have made some modifications to the original PicSOM system that affect mostly feature extraction and visualization parts of PicSOM. As an extra problem-specific knowledge we have segmentation masks for each defect image. This information is utilized in PicSOM so that feature extraction is only done for defect areas in each defect image.

Overview of the method

Interpretation of defect images is a demanding task even to an expert. The defect images concerned in this work contain surface defects, and they were taken from a real, online process. Currently we have two major database types: paper and metal surface defects. Both of these types contain several different defect classes (e.g. dark and light spots, holes, scratches, oli stains and so on) that are fuzzy and overlapping, so it is not possible to label defects unambiguously.

In the present work we have adopted the PicSOM system as our content-based image retrieval (CBIR) system and embedded the defect image databases into PicSOM. PicSOM has several features that make it a good choice for our purposes. The most important of these is the fact, that PicSOM can effectively combine search results of different features. This makes adding new features fast and efficient.

Features for defect characterization Several types of features can be used in PicSOM for image querying. These include features for color, shape, texture, and structure description of the image content. When considering defect images, there are two types of features that are of interest: shape features and internal structure features. Shape features are used to capture the essential shape information of defects in order to distinguish between differently shaped defects, e.g. spots and wrinkles. Internal structure features are used to characterize the gray level and textural structure of defects.

One of the advantages of PicSOM is its open architecture. This makes it simple to add new features to the system. Originally we used simple descriptors for shape, texture features based on the co-occurrence matrix, and the gray level histogram. We use the following set, which consists mainly of features from the MPEG-7 standard, with some additional features that we have developed for the shape description of surface defects.

Scalable Color descriptor is a 256-bin color histogram in HSV color space, which is encoded by a Haar transform.

Color Layout descriptor specifies a spatial distribution of colors. The image is divided into 8×8 blocks and the dominant colors are solved for each block in the YCbCr color system. Discrete Cosine Transform is applied to the dominant colors in each channel and the DCT coefficients are used as a descriptor.

Color Structure descriptor captures both color content and the structure of this content. It does this by means of a structuring element that is slid over the image. The numbers of positions where the element contains each particular color is recorded and used as a descriptor. As a result, the descriptor can differentiate between images that contain the same amount of a given color but the color is structured differently.

Edge Histogram descriptor represents the spatial distribution of five types of edges in 16 sub-images. The edge types are vertical, horizontal, 45 degree, 135 degree and non-directional, and they are calculated by using 2×2-sized edge detectors for the luminance of the pixels. A local edge histogram with five bins is generated for each sub-image, resulting in a total of 80 histogram bins.

Homogeneous Texture descriptor filters the image with a bank of orientation and scale tuned filters that are modeled using Gabor functions. The first and second moments of the energy in the frequency domain in the corresponding sub-bands are then used as the components of the texture descriptor.

Shape feature For shape description, we use our own problem-specific shape feature set that was developed for surface defect description. It consists of several simple descriptors calculated from a defect's contour. The descriptors are convexity, principal axis ratio, compactness, circular variance, elliptic variance, and angle.

Edge Co-occurrence Matrix was developed for enhanced shape description. It is similar to the Gray Level Co-occurrence Matrix, but instead of a gray level intensity image, it is computed from an edge image. In this respect it is related to the Edge Histogram, but the entire image is processed at once and eight edge directions at 45 degree intervals are detected. The result is an 8×8 matrix representing the frequencies of edge direction pairs that have a given displacement.

Contour Co-occurrence Matrix was developed in conjunction with the Edge Co-occurrence Matrix. It is computed from the chain-code representation of the contour of a defect. Each link in the chain is a step between two successive pixels on the contour, with eight possible link directions. The result is an 8×8 matrix representing the frequencies of link direction pairs that have a given displacement.

These features were found to work very well on classification experiments using a smaller, pre-classified data base. Similar or better performance is obtained when the Edge Co-occurrence Matrix and the Contour Co-occurrence Matrix are substituted for the Edge Histogram descriptor and the earlier shape feature set.

Experiments

The problem at hand is now the following one: Given a new defect or a set of defects, retrieve similar defects that might have appeared previously. The retrieval is based on shape and internal structure features, so there is no need for manual annotation or labeling. The largest defect database has almost 45000 defect images that were taken from a real, online process. The images have different kinds of defects, e.g. dark and light spots, holes, and wrinkles. They are automatically segmented beforehand so that each defect image has a gray level image and a binary segmentation mask that indicates defect areas in the image. The image database was provided by our industrial partner, ABB oy.

Two example queries in Figure 7.6 show that the system works quite well. Under the TS-SOMs are the images selected by the user (the so called query images), and at the

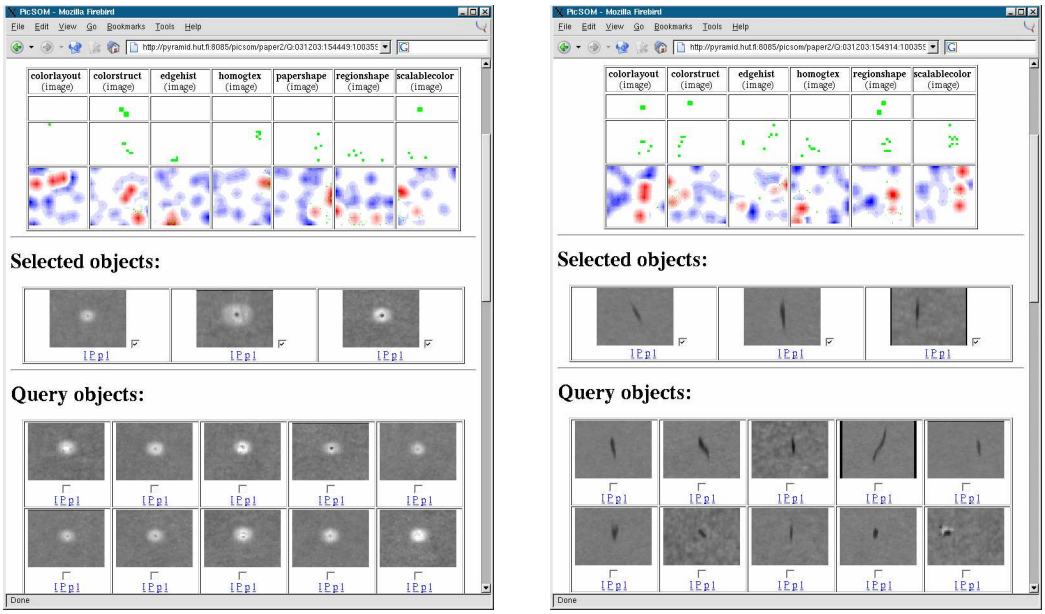

Figure 7.6: Example PicSOM queries.

bottom are the images returned by the PicSOM system. All returned images are visually similar to the query images. The system retains a similar level of success when queried with different types of defects. The true power comes from combining the maps. The PicSOM engine combines the various maps in a powerful manner, yielding good results.

Conclusions

In this project a noncommercial content-based image retrieval (CBIR) system called PicSOM is applied to retrieval of defect images. New feature extraction algorithms for shape and internal structure descriptions are implemented in the PicSOM system. The results of experiments with almost 45000 surface defect images show that the system works fast with good retrieval results.

References

[1] J. Iivarinen and J. Pakkanen, Content-Based Retrieval of Defect Images, In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, pp. 62–67, 2002.

[2] J. Pakkanen and J. Iivarinen, Content-based retrieval of surface defect images with MPEG-7 descriptors, In K. Tobin Jr. and F. Meriaudeau, editors, *Proceedings of Sixth International Conference on Quality Control by Artificial Vision*, Proc. SPIE 5132, pp. 201–208, 2003.

[3] R. Rautkorpi, J. Iivarinen, Shape-Based Co-occurrence Matrices for Defect Classification, In *Proceedings of 14th Scandinavian Conference on Image Analysis*, pp. 588–597, 2005.

7.3 Alterable volume flow in the use of input deformations for a massive Gaussian process smoothing

Ramūnas Girdziušas, Jorma Laaksonen

The use of input deformations with nonlinear regression models is relatively unpopular for a reason that they introduce another degree of nonlinearity into the model identification. However, in a massive regression with low-dimensional inputs (d=1,2,3) they have a visual representation and enhance the predictive power of simple stationary isotropic homogeneous models. Considering Gaussian process regression, input deformations are linked directly to the smoothing properties of the covariance functions:

$$\mathrm{Cov}\left[f(\mathbf{x}_i), f(\mathbf{x}_j)\right] \;\; \mapsto \;\; \mathrm{Cov}[f(\mathbf{x}_i - \mathbf{u}(\mathbf{x}_i)), f(\mathbf{x}_j - \mathbf{u}(\mathbf{x}_j))]. \tag{7.1}$$

A direct maximum likelihood estimation of the displacement field $\mathbf{u}(\mathbf{x})$ is hardly possible because it is unclear what smoothness priors to utilize and how to maintain the model space within the computationally feasible level.

We are investigating a likelihood-ascending nonparametric estimation of displacements which allow to alter the deformation volume and is computationally feasible. A suggestion is to utilize covariance functions of a univariate Brownian motion and extend them to a higher-dimensional case via an additive operator splitting [4]. Input deformations of such a model do not alter the structure of the inverse covariance matrix. It will remain tridiagonal [1] which makes the evaluation of the log-likelihood particularly easy. The Gaussian process log-likelihood $\mathcal{L}_u$ can then be maximized by embedding the displacement field into a simplified fluid flow [3]. This amounts to performing a non-gradient descent on $-\mathcal{L}_u + \mathbf{R}_{\lambda,\mu}$, where the Lamé functional $\mathbf{R}_{\lambda,\mu}$ is a positive definite quadratic form of the spatial derivatives of the displacements. However, it does not act as a regularization functional:

$$\mathcal{L}(\mathbf{u}_t + \mathbf{du}_t) \approx \mathcal{L}(\mathbf{u}_t) + \nabla_u \mathcal{L}(\mathbf{u}_t)^T \mathbf{du}_t, \tag{7.2}$$

$$\mathcal{R}_{\lambda,\mu}(\mathbf{u}_t + \mathbf{du}_t) \mapsto \mathcal{R}_{\lambda,\mu}(\mathbf{u}_t) + \mathbf{du}_t^T \nabla_u \nabla_u^T \mathcal{R}_{\lambda,\mu}(\mathbf{u}_t) \mathbf{du}_t \;, \tag{7.3}$$

The constants λ and μ define the compressibility of the flow. This approach has been tested in a synthetic problem of a bivariate regression where it is equally competitive with a robust but not visual regression techniques [3] and in the estimation of an optical flow, which does not reach the state of the art but is not dramatically far from it [2].

References

[1] R. Girdziušas and J. Laaksonen. Use of input deformations with Brownian motion filters for discontinuous regression. In *ICAPR*, volume 3686 of *LNCS*, p. 219–228. Springer, 2005.

[2] R. Girdziušas and J. Laaksonen. Optimal ratio of Lamé moduli with application to motion of Jupiter storms. In *SCIA*, vol. 3540 of *LNCS*, p. 1096–1106. Springer, 2005.

[3] R. Girdziušas and J. Laaksonen. Gaussian process regression with fluid hyperpriors. In *ICONIP*, vol. 3316 of *LNCS*, pages 567–572. Springer, 2004.

[4] B. Fischer and J. Modersitzki. *Inverse Problems, Image Analysis, and Medical Imaging*, volume 313 of *AMS Contemporary Mathematics*, chapter Fast Diffusion Registration, pages 117–129. 2002.

7.4 Stopping criteria for nonlinear diffusion filters

Ramūnas Girdziušas, Jorma Laaksonen

Nonlinear diffusion filters preserve discontinuities without requiring a prior knowledge of their approximate locations and reveal the optimal basis of a signal in an unsupervised manner [1]. Potential applications include a fast retrospective multiple change-point detection in nuclear magnetic response measurements of well log data, restoration of aerial images of buildings, fingerprint enhancement and optical flow.

We consider a nonlinear variation-diminishing diffusion of an image $y(\mathbf{x})$, defined by

$$\partial_t f = \nabla \cdot (\phi_\theta(||\nabla f||)\nabla f)\,, \quad s.t.\ \nabla f|_\Omega = 0, \tag{7.4}$$

with $f \equiv f_t(\mathbf{x})$, $f_0(\mathbf{x}) = y(\mathbf{x})$, $\mathbf{x} \in \Omega = [0,1]^d$, $d = 1, 2$. The univariate diffusivity function $\phi_\theta(z) \geq 0$ depends on a few parameters $\boldsymbol{\theta}$ and it is such that $\phi'_\theta(z)/(z)$ is continuous and decreasing.

Eq. (7.4) produces a scale-space of images and needs systematic criteria for the selection of the optimal stopping time. A possible remedy to this problem is to extend Eq. (7.4) in order to define a joint probability measure for the observations, diffusion outcome and the optimal stopping time, which could be accomplished within the framework of Itô diffusions. This turns out to be a formidable task.

Among a variety of heuristic stopping principles proposed in the last decades, one of the simplest and the most stable criterion stops the diffusion when the filtering outcome becomes uncorrelated with the noise estimate between the observations and a true signal [4]. Motivation for the decorrelation is based on the theory of a linear diffusion.

We have shown [1] that the decorrelation principle is a particular case of a maximum likelihood estimation in an additive white Gaussian noise when the assumption of a true signal is the Gaussian process defined by

$$F_t \sim \mathcal{GP}\left(0, \theta_0 \left(T_\theta - Id\right)^{-1}\right), \tag{7.5}$$

where θ_0 is a known constant, T_θ denotes the operator that maps an initial image to the diffusion outcome, it depends on the unknown diffusivity parameters $\boldsymbol{\theta}$ and the stopping time t. Such a model is useful because: (i) it explains why decorrelating criterion can be negative in nonlinear diffusion and gives its precise bounds, (ii) shows that the decorrelation principle overestimates the optimal stopping time, and (iii) provides better stopping criteria [2, 3].

References

[1] R. Girdziušas and J. Laaksonen. Gaussian processes of nonlinear diffusion filtering. In *IJCNN*, vol. 2, p. 1012–1017. IEEE, 2005.

[2] R. Girdziušas and J. Laaksonen. Jacobi alternative to Bayesian evidence maximization in diffusion filtering. In *ICANN*, vol. 3697 of *LNCS*, p. 247–252. Springer, 2005.

[3] R. Girdziušas and J. Laaksonen. Optimal stopping and constraints for diffusion models of signals with discontinuities. In *ECML*, vol. 3720 of *LNAI*, p. 576–583. Springer, 2005.

[4] P. Mrázek and M. Navara. Selection of optimal stopping time for nonlinear diffusion filtering. *Int. Journal of Computer Vision*, 52(2):189–203, 2003.