

IMAG Project

PRIMA-II

Perceiving Appearance for
Robust Real-time Computer Vision

Activity Report

James L. Crowley

1 December 2000

Contents

1. Appearance: A new theory of Visual Representation.....	1
2. Sampling, Indexing and Matching the Appearance Manifold.....	2
2.1 Gaussian receptive fields	2
2.2 Describing appearance with receptive fields.....	4
3. Appearance Manifolds for Robust Real Time Visual Perception.....	5
3.1 View-invariant object recognition using receptive field histograms	5
3.2 Indexing using histograms.	6
3.3 Recognizing patterns of activities.....	7
3.4 Robust probabilistic tracking of appearance.....	8
3.5 View invariant recognition using appearance manifolds.....	10
4. Current Problems	13
Publications of Project PRIMA II.	15
Doctoral Theses:.....	15
Book Chapters.....	15
Journal Publications.....	15
Reviewed papers in National and International Conferences.....	16
Reviewed Papers in Workshops.....	18
Bibliography.....	19

1. Appearance: A new theory of Visual Representation

During the period 1989 - 1995, the personnel of project PRIMA participated in a high-profile European Basic Research consortium named Vision as Process (VAP). The VAP consortium sought to investigate new models for control of perception for real time active vision. Project VAP met its goals and provided its demos [1]. However the project also demonstrated that current theories at the foundations of computer vision were inadequate.

The Project PRIMA II was formulated to address a failure at the foundation of computer vision made evident by VAP: The lack of an adequate theory for the representation of visual appearance.

The current practice in computer vision is to describe the contents of an image with one of three techniques: Region segmentation, interest point extraction, or edge detection. In all three cases the description processes are invented based on intuition and hunch. The descriptions that results are unstable and difficult to use for matching or interpretation.

In PRIMA II we have sought to develop a new approach for representing the information contained in images and image sequences. This approach provides a theoretical foundation for the design of operations for representing, recalling and matching image information. It makes possible the design of operations whose characteristics are mathematically predictable and whose calculations can be computed within real time processes. We believe that this approach provides a solid theoretical foundation for the science of computer vision. This approach is known as "appearance based" vision.

The appearance of an object, scene, or an event is the set of all possible images that may be observed. For example, the appearance for an object is the set of images which can be formed from all possible view directions and distances, under all possible lighting conditions. This space is captured by a "Plenoptic function" [2]. The domain of definition of the plenoptic function is the set of parameters for image formation. If we consider each image to be a form of feature vector, then the plenoptic function defines a manifold in the space of features. The key to appearance based vision is to provide an appropriate representation for this manifold.

Project PRIMA II has pioneered a new representation for visual appearance based on receptive fields. A receptive field is a local function defined over the domain of image position, time, and pixel color. For example, figure 1 shows the set of achromatic receptive fields using Gaussian derivatives developed by Vincent Colin de Verdiere.

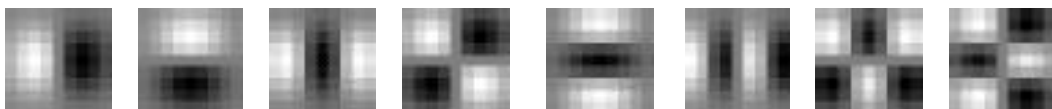


Figure 1. Achromatic receptive fields defined by the first three derivatives of a 2D Gaussian

2. Sampling, Indexing and Matching the Appearance Manifold

Receptive fields provide a basis for describing the local structure of signals [3]. In PRIMA II we have experimented with receptive fields defined over image position, time, and color. A projection of the appearance manifold onto a vector of receptive fields provides a vector of features. Such a vector, provides a concise description of the appearance within a local neighborhood. The vector of receptive field responses can be used as an address for rapid indexing into an appearance manifold. Such association can provide object identity, view direction and other relevant parameters in an operation which is equivalent to table look-up. Because the receptive field function is local, such association is robust to occlusions and scenes composed of multiple objects.

2.1 Gaussian receptive fields

The Hermite polynomial [4] of Gaussian derivatives has a variety of properties which make it ideally suited as a basis for receptive fields. The Gaussian is the function with optimal compactness in space and frequency [5], as measured by the product of the second moments. Thus the spectrum of the Gaussian is minimally deformed when the Gaussian is truncated to a finite spatial extent. The Gaussian exhibits logarithmic scale equivariance, which means that structures will translate in a log scale-space when changed in scale [6], [7]. As a result, it is possible to define scale equivariant detection measures which can be applied to log-scale spaces defined using a Laplacian of a Gaussian or the gradient magnitude of a Gaussian [8]. Such measures provide a means to easily determine the appropriate scale for description at each image point, in a way that varies uniformly with changes in scale.

The Gaussian function has the form

$$G(x, \sigma) = e^{-\frac{x^2}{2\sigma^2}}$$

where x is the spatial variable, and σ a scale variable that specifies the second moment of the function. The first three derivatives of the Gaussian are:

$$G_x(x, \sigma) = -\frac{x}{\sigma^2} G(x, \sigma)$$

$$G_{xx}(x, \sigma) = \frac{x^2 - \sigma^2}{\sigma^4} G(x, \sigma)$$

$$G_{xxx}(x, \sigma) = -\frac{x^3 - 3x\sigma^2}{\sigma^6} G(x, \sigma)$$

Receptive fields may be obtained by sampling these function over a finite interval. They may also be designed as recursive filters [9]. Even derivatives will respond to symmetric structures such as spots and bars while odd derivatives will respond to a-symmetric structures, such as edges. In two dimensions, the Gaussian is the unique function which is both separable and circularly symmetric. The Gaussian derivatives in 2D are steerable [10], making efficient computations possible. Oriented derivatives can be defined as a weighted sum of separable components.

2.2.1 Color opponent space

Color cameras commonly return images observed through three spectral filters R, G, B. These signals may be projected onto a Cartesian representation composed of a Luminance part L, and two orthogonal chrominance parts, C₁ and C₂.

$$\begin{array}{rcl} L & 0.33 & 0.33 & 0.33 & R \\ C_1 & = & 0.5 & -0.5 & 0 & G \\ C_2 & & 0.5 & 0.5 & -1 & B \end{array}$$

The luminance component varies with the angle between the surface normal and the illumination, and is thus an indicator of 3D shape. The chrominance components are determined by the surface pigment of objects, and are thus a signature for object recognition. The components C₁ and C₂ encode the chromatic information in a Cartesian representation which is orthogonal to the luminance. Computing derivatives over these two chrominance components leads to color opponent receptive fields which are suited for detecting adjacent associations of colors.

2.2.2 Invariance to rotation and scale

At each image position, it is possible to determine an intrinsic scale, s_i , and intrinsic orientation, θ_i . Intrinsic scale is determined by the local maxima in the Laplacian across scales. For each position (x, y) in an image, there is a discrete set of scales for which the inner product of the Laplacian with an image is a local maximum. This set may be discovered by searching for local maxima across scales.

$$s_i(x, y) = \text{Arg Max} \{ \langle A(x, y), \nabla^2 G(x, y, s) \rangle \}$$

Local maxima in the scale profile provide a reference for locally normalizing scale of the receptive fields [11]. The value s_i is perfectly equivariant to scale. Such normalization provides both robustness to changes in distance, and adaptation of the receptive field to the most appropriate scale for describing the appearance at each image position.

Orientation invariance is obtained by normalizing the orientation to the angle measured by the gradient. Whenever the gradient is not close to zero, it is possible to estimate an intrinsic orientation using the arc-tangent of the ratio of first derivatives. For a gradient calculated using Gaussian derivatives, G_x and G_y , this operation is

$$\theta(x, y) = \text{Atan2}\{\langle A(x, y), G_y \rangle, \langle A(x, y), G_x \rangle\}$$

The ATAN2 function avoids division by zero for horizontal edges and returns a value in the range $[-\pi, \pi]$. Oriented Gaussian derivatives may be synthesized using the steerability property [10]. Let $G_1(x, y, \theta)$ indicate a first derivative oriented at an angle of θ . Oriented receptive fields may be computed from derivatives in the Cardinal directions.

$$G_1 = \cos(\theta) G_x + \sin(\theta) G_y$$

$$G_2 = \cos(\theta)^2 G_{xx} + 2 \cos(\theta) \sin(\theta) G_{xy} + \sin(\theta)^2 G_{yy}$$

The steerable chromatic basis is a set of color opponent filters as shown in figure 2. Permutations of RGB lead to different opponent color spaces. The choice of the most appropriate space depends on the chromatic composition of the scene.



Figure 2. Chromatic Gaussian derivatives $G_x^L, G^{C1}, G^{C2}, G_x^{C1}, G_x^{C2}, G_{xx}^L, G_{xy}^L, G_{yy}^L$

D. Hall has developed a vector of eight achromatic Gaussian receptive fields, G_C , normalized to the local scale and orientation [12].

$$G_C = (G^{C1}, G^{C2}, G_x^L, G_x^{C1}, G_x^{C2}, G_{xx}^L, G_{xy}^L, G_{yy}^L)$$

This set has been found to have excellent properties for object recognition, defect detection and image data base indexing.

2.2 Describing appearance with receptive fields.

Let $A(x, y, P)$ represent the appearance of an object or scene at image position x, y and image formation parameters P . Appearance is described by projection onto a vector of receptive fields $G(\theta, \sigma)$ at the intrinsic scale and orientation, (θ, σ) ,

$$M(P) = \langle A(x, y, P), G(\theta, \sigma) \rangle$$

where $\langle \cdot, \cdot \rangle$ is the inner product operation. The vector M defines a point in a space defined by G . An image of an object is a grid of such points. Adjacent points on this grid tend to be adjacent in appearance space, and can be seen as sample points on a surface. An example is shown in figure 5, which shows the grid of responses for three receptive fields G^{C1} , G^{C2} , G_x^L , obtained from the region of the pen image illustrated by the red square. Similarity of local appearance makes it possible use local appearance as an index, or address, to recall manifolds with similar appearance. Regions of images with similar appearance may be associated using distance in the local appearance space. A vector of responses may be used to determine a list of manifolds which pass within a volume of radius " " .



Figure 5. A pen and the surface produced by projection onto three receptive fields G^{C1} , G^{C2} , G_x^L

A sphere of view points around an object defines a 4 dimensional grid in appearance. Connecting the grids across adjacent images gives a sampling of the appearance manifold which can be used for view invariant object recognition. A key problem has been to organize and represent such manifolds to allow fast and robust indexing and matching for visual perception.

3. Appearance Manifolds for Robust Real Time Visual Perception

The PRIMA-II project is currently exploring two different approaches to representing and matching the manifolds of appearance. One approach involves representing the structure of the manifold an M dimensional grid in an N dimensional feature space of receptive field responses. The second approach involves capturing the joint statistics of receptive field responses in multi-dimensional histograms.

3.1 View-invariant object recognition using receptive field histograms

The representation and matching of appearance with multi-dimensional receptive field histograms has been developed in the thesis of Bernt Schiele. Schiele represents the projection of an image neighborhood onto a receptive field as a coarsely quantified random variable. Given a sufficiently large number of samples of a discretely quantified random variable, a probability density function may be approximated using a histogram. For joint statistics, the number of required samples grows exponentially with the number of dimensions. However, a stream of video data contains a very large number of pixels.

Schiele experimentally compared a variety of recognition techniques using multidimensional receptive field histograms. A technique based on the use of Bayes rule was found to give clearly superior recognition rates. For his experiments, Schiele employed data bases of hundreds of images of objects

observed from different view points. These images were described using a vector of 8 achromatic Gaussian receptive field which had been normalized to the local image orientation.

Consider a set of N_O images of an object "o". Let $h_o(M)$ represent the histogram of joint statistics of the receptive field responses for pixels from the object. Then the probability of obtaining M given a pixel of the object is

$$p(M|o) = \frac{1}{N_O} h_o(M)$$

Let o be one of K objects, each represented by N_k pixels. The total number of pixels N_{tot} and the histogram of all responses $h_{tot}(M)$ are

$$N_{tot} = \sum_{k=1}^K N_k \quad h_{tot}(M) = \sum_{k=1}^K h_k(M)$$

The probability of encountering the receptive field response M in the entire image set is

$$p(M) = \frac{1}{N_{tot}} h_{tot}(M)$$

The probability of encountering a pixel from object o is

$$p(o) = \frac{N_o}{N_{tot}}$$

By Bayes rule, the probability of observing object o given a receptive field response M is a ratio of histograms, which may be captured in a single lookup table,

$$T_o(M) = p(o|M) = \frac{p(M|o)p(o)}{p(M)} = \frac{\frac{1}{N_o}h_o(M) \frac{N_o}{N_{tot}}}{\frac{1}{N_{tot}}h_{tot}(M)} = \frac{h_o(M)}{h_{tot}(M)}$$

Schiele has shown that histograms of receptive field responses provide a robust and fast method to recall and recognize objects. Duffy has recently demonstrated a variation of this technique for visual inspection using chromatic receptive fields.

3.2 Indexing using histograms.

Chomat and Pellison have developed an extension of the quad-tree technique to represent histograms with 12 or more dimensions. A dichotic tree is used to represent up to 2^N potential branches corresponding to filled cells. The value N represents the logarithm of the quantification of the receptive field values. As data arrives at a cell, the cell is sub-divided by 2 along each dimension. Only filled cells are encoded, which makes the representation appropriate for sparse histograms.

Among the $2N$ resulting new cells, the filled cells are sub-divided themselves until the final resolution of N bits for the receptive field projection. This algorithm allows the computation and the storage of high dimensional histograms which are quite sparse.

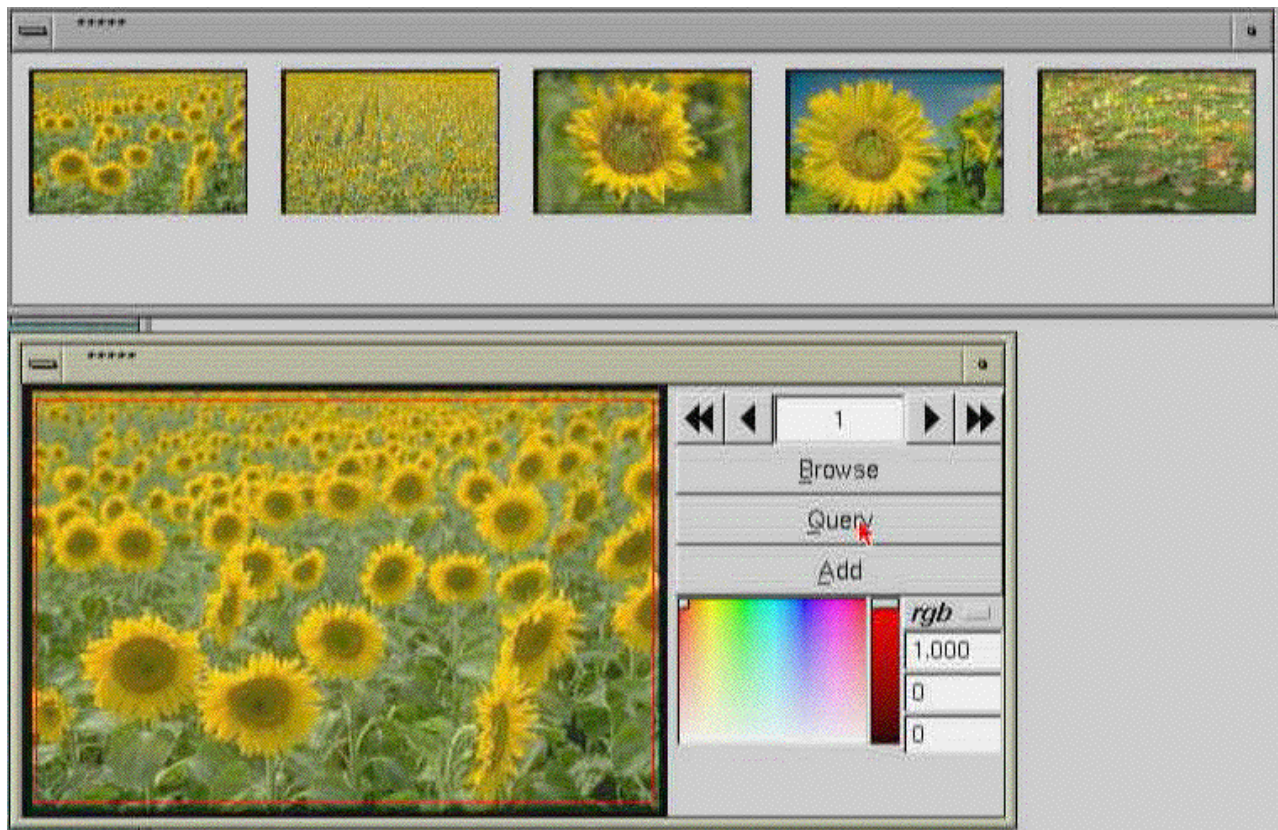


Figure 6. Image data based indexing using receptive fields.

Pelisson has adapted this technique to indexing into large data-bases of images. Pelisson's system uses the chromatic receptive fields of figure 6 to index into a data base of 40 000 images using queries based on samples of appearance.

3.3 Recognizing patterns of activities

Spatio-temporal appearance can be described and recognized by projection onto spatio-temporal receptive fields. Chomat has defined a family of spatio-temporal receptive fields for recognizing patterns of activity based on Gabor filters. Ratios of these receptive fields are used to describe local motion patterns based on spatio-temporal frequency independent of surface texture. Patterns of activity are recognized using a 12 dimensional receptive field histogram. Chomat has used this technique to fine probabilistic sensors for activity patterns. Such sensors may be used to compute the probability of an activity pattern in an image sequence for any size region of the image including individual pixels.

Computing the probability of an activity from a ratio of histograms for individual pixels is known as "back-projecting" the probabilities. Back-projection produces an image in which each pixel represent

the probability of the event to be detected. An example is shown below in figure 7. In this example, the spatio-temporal activity patterns have been learned for three species of gold-fish using ratios of Gabor filters to provide 12 spatio-temporal receptive fields. The three images on the right show the probability that each pixel from the sequence on the left is a member of each of the three species. White represents a high probability. This techniques is used to recognize human activity in an office environment.

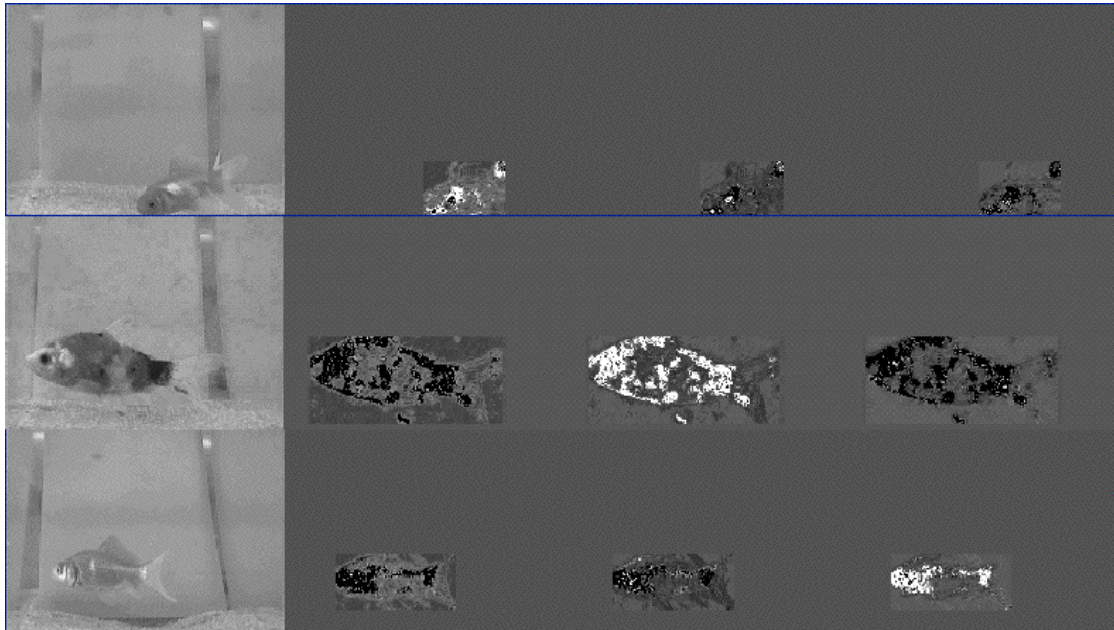


Figure 7. Recognizing Goldfish from their motion pattern.

The probabilistic sensor of Chomat is one several examples of pixel level detection functions that can be developed from receptive field vectors. Karl Schwerdt has recently developed a robust probabilistic tracking techniques for tracking objects detected with such functions.

3.4 Robust probabilistic tracking of appearance

The reflectance function of human skin may be modeled as a sum of a Lambertian and a specular reflectance function. In most cases the Lambertian component dominates. For a Lambertian surface, the intensity of reflected light varies as a function of the cosine of the angle between the surface normal and illumination. Because the face is a highly curved surface, the observed intensity of a face exhibits strong variations. These variations may be removed by dividing the three components of a color pixel, (R, G, B) by the intensity. This gives an intensity-normalized color vector, with two components, (r, g).

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B}$$

The intensity-normalized pixels from a region of an image known to contain skin can be used to define a two dimensional histogram, $h_{\text{skin}}(r, g)$, of skin color. A second histogram, $h_{\text{tot}}(r, g)$ can be

made from all of the pixels of the same image. These two histograms make it possible to apply Bayes rule to each pixel of an image to obtain the probability that a given pixel is skin.

$$p(\text{skin} | r, g) \quad h_{\text{ratio}}(r, g) = \frac{h_{\text{skin}}(r, g)}{h_{\text{total}}(r, g)}$$

This histogram ratio is trivial to build and may be renewed whenever an independent source has detected the face in the image.

In order to track faces we must group skin pixels into a region. Schwerdt has recently developed a novel robust tracking method. Let $P_{\text{skin}}(i, j)$ represent the probability map of skin for each color pixel $(r(i, j), g(i, j))$ at position (i, j) .

$$p_{\text{skin}}(i, j) = p(\text{skin} | r(i, j), g(i, j))$$

The center of gravity or first moment of the probability map gives the position and spatial extent of the skin colored region.

$$\mu = \begin{matrix} \mu_i \\ \mu_j \end{matrix} \quad \mathbf{C} = \begin{matrix} i^2 & ij \\ ij & j^2 \end{matrix}$$

An estimate from a prior image can be used to calculate the probability that a pixel belongs to a face in the current image.

$$p_t(\text{face}(i, j)) = p_{\text{skin}}(i, j) p_{t-1}(\text{face}(i, j))$$

The probability of the face can be estimated from a Gaussian function evaluated using the previous center of gravity and covariance. This processes recursively tracks detected regions while rejecting outliers.

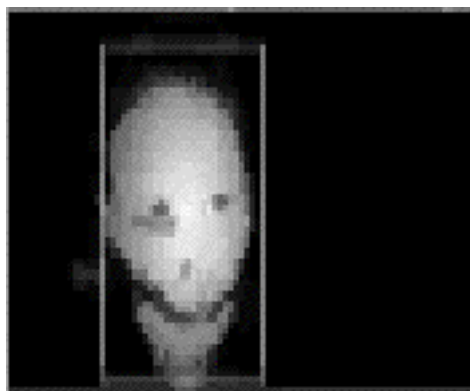


Figure 8 Probability of skin tracked using the robust probabilistic tracker of Schwerdt.

The robust tracking technique of Schwerdt has been used with variety of pixel level detection functions.

3.5 View invariant recognition using appearance manifolds

A structural representation of the appearance manifold makes it possible to recognize objects from arbitrary view-points based on observation of a single image neighborhood. Such an approach was originally pioneered in the doctoral thesis of Vincent Colin de Verdiere using achromatic receptive fields. Daniela Hall is currently extending this technique to observe and recognizing the manipulation of physical icons or "Phycons".

A view-sphere for an object is the set of viewing positions on a sphere of radius, R , centered on and enclosing the object. Positions on the view-sphere are parameterized by angles θ and ϕ representing latitude and longitude. An image projection axis runs from the center of the sphere through each point on the view sphere to a retina located at a distance, F , outside the sphere. The orientation of the retina around the projection axis is labeled by the angle ψ . Thus the vector (R, θ, ϕ, ψ) provides a spherical coordinate representation for view-points of an object placed in the center of a view-sphere.

To provide data for experiments, D. Hall have used a robot arm to construct images of an object at regularly spaced view-points. A "Geodesic Dome" [13] is used to define a "geodesic sampling function" for the view sphere. Sampled points on the geodesic dome are obtained by embedding an icosahedron within the unit sphere. The viewpoints on the vertices of the original icosahedron each have 5 equidistant neighbors, while other vertices have 6 equidistant neighbors. For these experiments she has constructed data sets of images at a distance of $R/10$, to give 91 images for a hemisphere, as shown in figure 9.



Figure 9. A geodesic sample function for the view sphere defined by a Geodesic dome with a spacing of $R/10$. The half sphere contains 91 images. On the right a set of 90 images sampled with a geodesic function.

The indexing structure is a form of tree in which each dimension of the appearance vector is represented by a branch. Each branch is organized into a set of buckets composed of, at most, N samples. Each bucket points to a list of possible branches based on the next dimension in the appearance space. A form of branch and bound is used to assemble lists of hypothesis with similar appearance.

Let o be the object, (i,j) an image position, (m, n) a position on the view sphere and d is the distance in local appearance space. The appearance manifold is represented by four structures.

$M_O(i, j, m, n)$: The local appearance vectors

$N_O(i, j, m, n)$: The intrinsic orientation and scale.

$D_O(a, i, j, m, n)$: the disparity tables which provides connectivity to each of the neighboring images, $a= 1, 2, 3, 4, 5, 6$.

$T()$: The indexing structure which predict a lists of hypotheses of the form (o, i, j, m, n, d) for an appearance vector M .

Given a vector, M , the indexing function returns a sorted list of hypotheses for the object identity, view-point and image position. The table $M_O(i, j, m, n)$ is necessary for selecting between competing hypotheses in a prediction-verification algorithm. The indexing structure $T()$ provides a list of likely manifolds as hypotheses for matching.

The indexing structure is a form of tree in which each dimension of the appearance vector is represented by a branch. Each branch is organized into a set of buckets composed of, at most, N samples. Each bucket points to a list of possible branches based on the next dimension in the appearance space. A form of branch and bound is used to assemble lists of hypothesis with similar appearance. For each volume of a manifold which passes within a distance of a distance of M in appearance space, the nearest point in that region is provided by a hypothesis encoded as a list H .

$H = (o, i, j, m, n, d)$.

The hypothesis states that for object o , the image position (i, j) in view point (m, n) has distance of d from M in appearance space. The indexing function returns a list of k hypotheses, $\{H_k\}$, sorted on distance.

The appearance manifold makes it possible to recognize objects from arbitrary view-points based on observation of a single image neighborhood. However, not all image neighborhoods provide unambiguous recognition. Some local appearance patterns such as straight edge lines and corners occur widely and can not be used to unambiguously recognize objects. Still other neighborhoods, such as image highlights, are completely unrecognizable. We classify image neighborhoods as belonging to one of three "types" according to their ability to discriminate objects.

Type 1: Image neighborhoods for which the indexing structure returns a small number (say < 10) of manifolds within a given distance.

Type 2: Image neighborhoods for which the indexing structure returns a large number (say 10) of manifolds within a given distance.

Type 3: Image neighborhoods which are unmatchable, either because of a lack of energy (constant regions) or because no manifold is found (for example, specularities).

Constructing appearance manifolds requires a number of design parameters. One important design parameter is the density with which the view sphere should be sampled in constructing the appearance manifold. An initial impression might be that the view sphere should be sampled as densely as possible. However, experience indicates that an overly dense sampling of the view sphere degrades the recognition process by encumbering the matching with a large number of very similar views. To illustrate this, appearance manifolds were constructed from the eight images of ordinary desktop objects shown in figure 10. This object set includes some interesting challenges for view invariant recognition. For example, the scissors, tape and pill-box contain specular regions. The protractor is partially transparent. The set contains both uniform and textured regions as well as curved and flat surfaces.

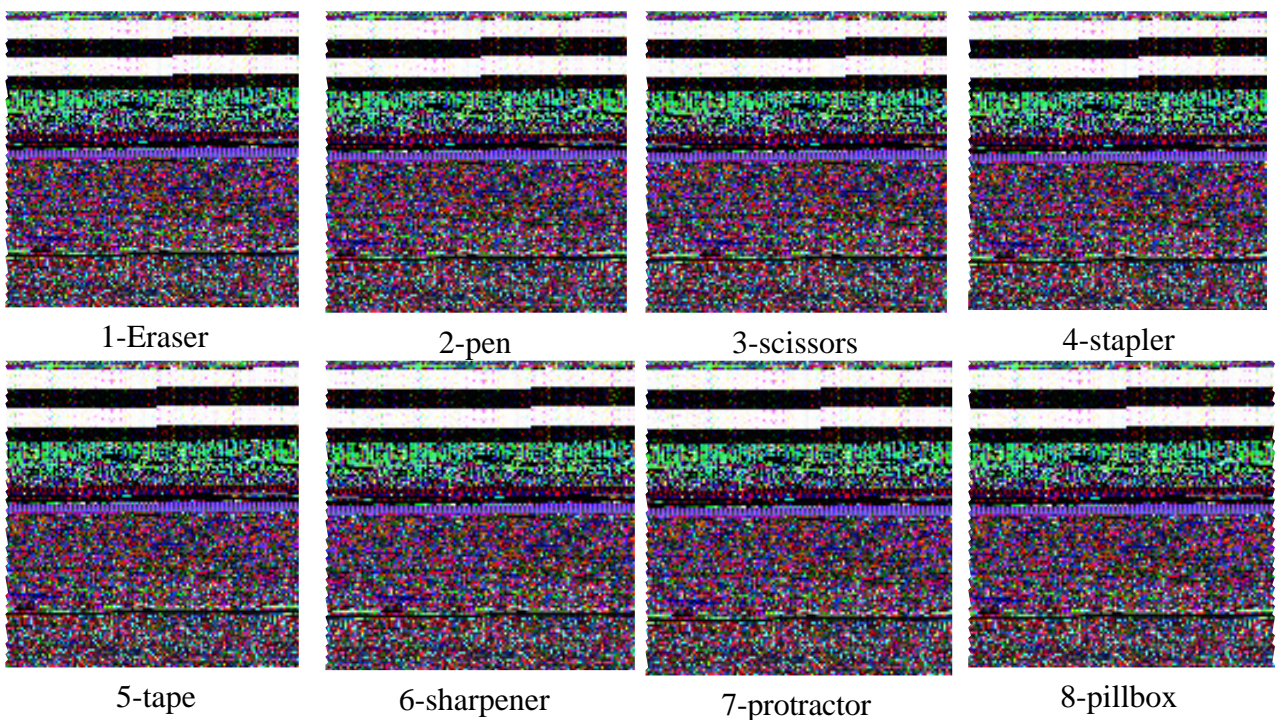
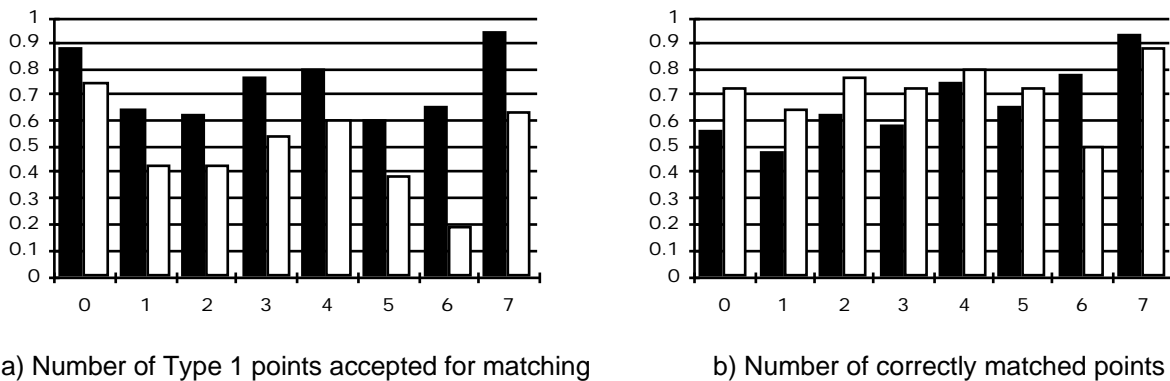


Figure 10. A set of 8 test objects used for experiments with chromatic receptive fields.

Two data sets were constructed for each of the eight objects. The first data set corresponds to the 91 vertices on a geodesic view sphere sampled at $\pi/10$. This data set was used to build a local appearance manifold for each object. The second data set was taken at the center of each triangular facet of the geodesic. To investigate the influence of the sample density of the view sphere, we formed appearance manifolds using 6 images with distance $\pi/3$ from the first data set and 26 images with distance $\pi/5$ for the second data set. Images of the view sphere which are not used for constructing the manifold were taken as test images. Each point of each test image was matched to the

8 manifolds. The percentage of points which were found to provide unique matches (Type 1) is shown in part a of figure 11. Part b of figure 11 shows the percentage of type 1 points for which the first hypothesis gave the correct object and view angle. A view angle is correct if it is the closest view angle include in the manifold.



a) Number of Type 1 points accepted for matching b) Number of correctly matched points
Figure 11. View Invariant Object Recognition. Comparison of matching with manifolds constructed from 6 images (black) and 26 images (white). Top is number of points which give a unique match. Bottom is number of points for which closest manifold belongs to correct view angle.

The results shown in figure 11 show that it is possible to use the appearance manifolds for view invariant object recognition using individual neighborhoods. They also show that as the density of the samples on the view sphere increases, robustness to changes in view direction increases, while the number of points judged to be unreliable also increases. This is because adjacent images from the view sphere generate competing hypotheses. As the number of competing hypotheses grows, the number of unique (type 1) matches decreases.

Sampling the view sphere at 22° between images provides only very coarse information about view angle. However, increasing angular discrimination requires increasing the sample density of the view sphere which, in turn, reduces the number of samples which provide unique indexing. Our conclusion is that improving the precision with which view angle can be estimated requires using structural constraints. We are currently working on such an algorithm using a form of prediction-verification.

4. Current Problems

As with an successful investigation, progress uncovers more questions than answers. A number of interesting questions have recently been opened for exploration. For example, it appears possible to extend the indexing technique of Pelisson from static images to image sequences. Chomat's activity recognition labels individual pixels. Questions remain about how to assemble regions of elemental activities into recognition of actions. We are currently exploring means to automatically acquire appearance manifolds by observing objects. A new line of research involves the automatic acquisition of recognition procedures based on appearance manifolds.

Publications of Project PRIMA II.

Doctoral Theses:

1. Olivier CHOMAT, "Caracterisation d'elements d'activites par la statistique de champs receptifs", Thèse doctorale de l'I.N.P.G., 19 septembre 2000. (avec félicitations du jury)
2. Jérôme MARTIN, "Reconnaissance de Gestes en Vision par Ordinateur", Thèse doctorale de l'I.N.P.G., 13 July 2000.
3. Vincent COLIN DE VERDIERE, "Représentation et Reconnaissance d'Objets par Champs Réceptifs", Thèse doctorale de l'I.N.P.G., 10 décembre 1999. (avec félicitations du jury)
4. Francois BERARD, "Vision par Ordinateur pour l'Interaction Fortement Couplée", Thèse doctorale de l'IU.J.F, 30 novembre 1999. (avec félicitations du jury) (Co-direction avec J. Coutaz)
5. Frank WALLNER, "Position Estimation for a Mobile Robot From Principal Components of Laser Rance Data", Doctoral Thesis of the I.N.P.G., 30 October 1997.
6. Bernt SCHIELE, "Reconnaissance d'Objets utilisant of Histogrammes Multi-Dimensionels of Champs Receptifs", Doctoral Thesis of the I.N.P.G., July 1997.
7. Bruno ZOPPIS, "Outils pour l'Intégration et le Contrôle en Vision et Robotique Mobile", Doctoral Thesis of the I.N.P.G., 22 June 1997
8. Stephen JONES, "Réalisation Robuste de Taches Concurrentes", Doctoral Thesis of the I.N.P.G., 26 May 1997.

Book Chapters

1. James L. Crowley, Mathematical Foundations of Navigation and Perception for an Autonomous Mobile Robot, in Reasoning and Uncertainty, edited by L. Dorst, Springer Verlag, 1997.

Journal Publications

1. D. Hall, J. L. Crowley and V. Colin de Verdière, "View Invariant Object Recognition using Coloured Receptive Fields", Machine Graphics and Vision, Vol 9. pp 341-352, No. 2. June 2000.
2. V. Colin de Verdière et J. L. Crowley, "Local Appearance Space for Recognition of Navigation Landmarks", Robotics and Autonomous Systems, Vol 31, Nos 1-2, pp 61-70, April 2000.
3. J. L. Crowley, J. Coutaz and F. Berard, "Things that See: Machine Perception for Human Computer Interaction", Communications of the A.C.M., Vol 43, No. 3, pp 54-64, March 2000
4. B.Schiele and J. L. Crowley, "Recognition without Correspondence using Multidimensional Receptive Field Histograms", International Journal of Computer Vision, 36(1), pp 31-50, Jan. 2000

5. J. L. Crowley, F. Wallner and B. Schiele, "Position Estimation Using Principal Components of Range Data", Robotics and Autonomous Systems, Vol 23, no 4, pp 267-276, 1998.
6. B. Schiele and J. L. Crowley, "Transformation of Object Recognition and its Application to Viewpoint Planning.", Robotics and Autonomous Systems, Vol 21, No 1, July 1997.
7. J. L. Crowley, "Vision for Man machine interaction", Robotics and Autonomous Systems, Vol 19, Nos 3-4, p 347-359, avril 1997.
8. P. Reignier, V. Hansen and J. L. Crowley, "Incremental Supervised Learning for Mobile Robot Reactive Systems", Robotics and Autonomous Systems, Vol 19, Nos 3-4, p 347-359, avril 1997.

Reviewed papers in National and International Conferences

1. K. Schwerdt, D. Hall, J. L. Crowley, "Visual Recognition of Emotional States", The Third International Conference on Multimodal Interfaces, ICMI-2000, Beijing China, 14-16 October 2000.
2. C. Le Gal, A E Ozcan, K Schwerdt, J L. Crowley, "A Sound MagicBoard", The Third International Conference on Multimodal Interfaces, ICMI-2000, Beijing China, 14-16 October 2000.
3. D. Hall, J. L. Crowley, "Estimating the Pose of Phicons for Human Computer Interaction", The Third International Conference on Multimodal Interfaces, ICMI-2000, Beijing China, 14-16 October 2000.
4. N. Duffy and J. L. Crowley, "Object Detection Using Colour", ICPR 2000, Barcelona september 2000.
5. O. Chomat, V. Colin de Verdiere, D. Hall and J. L. Crowley, "Local Scale Selection for Gaussian Based Description Techniques", 6th European Conference on Computer Vision, Springer Verlag, Dublin, pp 117-133, June 2000.
6. O. Chomat, J. Martin, and J. L. Crowley, "A Probabilistic Sensor for the Perception and Recognition of Activities", 6th European Conference on Computer Vision, Springer Verlag, Dublin, pp 487-506, June 2000.
7. D. Hall, V. Colin de Verdiere and J. L. Crowley, "Object Recognition using Coloured Receptive Field", 6th European Conference on Computer Vision, Springer Verlag, Dublin, pp 164-178, June 2000.
8. O. Chomat and J. L. Crowley, "A Probabilistic Sensor for the Perception of Activities", 4th IEEE International Conference on Automatic Face and Gesture Recognition", Grenoble, France, March 2000.
9. K. Schwerdt et J. L. Crowley, "Robust Face Tracking using Color", 4th IEEE International Conference on Automatic Face and Gesture Recognition", Grenoble, France, March 2000.
10. J. B. Durand et J. Martin, "Reconnaissance automatique de gestes d'Écriture par modèles de Markov cachés : étude de paramètres des modèles", in 12eme Congres de Reconnaissance des Formes et Intelligence Artificielle (RFIA'2000), February, 2000.

11. D. Hall and J. L. Crowley, "Tracking Fingers and Hands with a Rigid Contour Model in an Augmented Reality", MANSE 99: Managing Interactions in Smart Environments, ISBN 1-85233-228-X, Springer Verlag, pp 34-43, Dublin Dec 1999.
12. O. Chomat and J. L. Crowley, "Probabilistic Recognition of Activity using Local Appearance", IEEE Conference on Computer Vision and Pattern Recognition, CVPR 99, Fort Collins, June 1999.
13. J. Coutaz, F. Berard, E. Carraux and J. L. Crowley, "CoMedi: Using Computer Vision to Support Awareness and Privacy in Mediaspaces", ACM Conference on Computer Human Interaction, CHI 99, Pittsburgh, May 1999.
14. W. Vieux, K. Schwerdt, and J. L. Crowley, "Face-tracking and coding for video compression", ICVS 99, International Conference on Vision Systems, Springer LNCS Series, Jan 1999.
15. J. Coutaz, F. Berard, E. Carraux, et J. L. Crowley, "Early experiences with the mediaspace CoMedi", European Conference on Computer Human Interaction, EHCI98, Crete, Sept. 1998.
16. V. Colin of Verdiere and J. L. Crowley, "Visual Recognition Using Local Appearance", ECCV '98, Friburg, June, 1998.
17. G. D. Finlayson, B. Schiele and J. L. Crowley, "Comprehensive Color Image Normalisation", ECCV '98, Friburg, June, 1998.
18. James L. Crowley, Frank Wallner and Bernt Schiele, "Position Estimation Using Principal Components of Range Data", 1998 IEEE International Conference on Robotics and Automation, Leuven, May, 1998.
19. G. D. Finlayson, B. Schiele and J.L. Crowley, "Using colour for image indexing", in The Challenge of Image Retrieval, April 1998.
20. J. Martin and J. L. Crowley, "Active Hand Tracking", Third International IEEE Conference on Automatic Face and Gesture Recognition, Nara, Japan, April 14-16, 1998.
21. V. Colin of Verdière et J. L. Crowley, "Reconnaissance d'Objets par Apparence Locale", RFIA '98, Clermont-Ferrand, jan 1998.
22. A. Guarda et A. Lux, "Apprentissage of Détecteurs Visuels", RFIA '98, Clermont-Ferrand, jan 1998.
23. B. Schiele and J. L. Crowley, "Transinformation for Active Object Recognition", In ICCV'98, International Conference on Computer Vision, Bombay, India, jan 1998
24. J. L. Crowley, F. Wallner and B. Schiele, "Position Estimation Using Principal Components of Range Data", Intelligent Autonomous Robotics and Processes, Genoa, It, oct. 1997.
25. J. Martin et J. L. Crowley, "An Appearance-Based Approach to Gesture-Recognition", 9th International Conference on Image Analysis and Processing, Florence, Italy, September 1997
26. S. D. Jones, C. Anderssen et J. L. Crowley, "Appearance Based Processes for Visual Navigation", IROS '97, IEEE International Conference on Intelligent Robots and Systems, Grenoble, Sept. 1997.
27. J. Coutaz, J. L. Crowley, F Berard and D. Salber, "EigenSpace Coding as a Means to Support Privact in Computer Mediated Communication", Interact 97, Sydney, July 97.

28. B.Schiele and J. L. Crowley, "The concept of Visual Classes for Object Classification.", Scandinavian Conference on Image Analysis, SCIA '97, Lappeenranta, June 1997.
29. J. L. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications", IEEE Conference on Computer Vision and Pattern Recognition, CVPR '97, St. Juan, Puerto Rico, June 1997.
30. B. Schiele and J. L. Crowley, "The concept of Visual Classes for Object Classification.", Scandinavian Conference on Image Analysis, SCIA '97, Lappeenranta, June 1997.

Reviewed Papers in Workshops

1. J. L. Crowley and K. Schwerdt, "Robust Tracking and Compression for Video Communication", IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real Time Systems, Corfu, September 1999.
2. O. Chomat, V. Colin de Verdière and J. L. Crowley, " Recognizing Goldfish? or Local Scale Selection for Recognition Technique", Symposium on Intelligent Robotics Systems, SIRS-99, Coimbra, pp 197-206, July 1999
3. E. Backman V. Colin de Verdière et J. L. Crowley, "Dense Stereo Matching using Local Appearance", Symposium on Intelligent Robotics Systems, SIRS-99, Coimbra, pp13-22, July 1999.
4. D. Hall, C. Le Gal, J. Martin, O. Chomat, T. Kapuscinski, and J. L. Crowley, "MagicBoard: A Contribution to an Intelligent Office Environment", Symposium on Intelligent Robotics Systems, SIRS-99, Coimbra, pp215-224, July 1999.
5. F. Pourraz and J. L. Crowley, "Continuity Properties of the Appearance Manifold for Mobile Robot Position Estimation", IEEE Workshop on Visual Navigation, Fort Collins, June 1999.
6. F. Pourraz and J. L. Crowley, "Continuity Properties of the Appearance Manifold for Mobile Robot Position Estimation", Symposium on Intelligent Robotics Systems, SIRS-98, Edinburgh, July 1998.
7. J. L. Crowley, F. Wallner, and B. Schiele, "Position Estimation Using Principal Components of Range Data", IAR Annual Meeting, Duisberg, nov 1997.
8. M. Kroschel, M. S. Mekhaïel, J. L. Crowley and F. Bérard, "Fast Face Localisation System Using a Statistical Color Model", IAR Annual Meeting, Duisberg, nov 1997.
9. J. L. Crowley and J. Martin, "Visual Processes for Tracking and Recognition fo Hand Gestures", International Workshop on Perceptual User Interfaces, Banf, Ca, oct. 1997.
10. F. Wallner, J. L. Crowley et B. Schiele, "Position Estimation Using Principal Components of Range Data", SIRS '97, Stockholm, July 1997.
11. A. Lux, "The RAVI environment", SIRS '97, Stockholm, July 1997.

Bibliography

- [1] J. L. Crowley and H. I Christensen, Vision as Process, Springer Verlag, Heidelberg, 1994.
- [2] T. Adelson and J. Bergen, "The plenoptic function and the elements of early vision", in M. Landy and J. A Movsohns, eds., Computational Models of Visual Processing, MIT Press, Cambridge, 1991.
- [3] J.J. Koenderink, A.J. van Doorn, "Representation of Local Geometry in the Visual System", Biological Cybernetics, pages 367-375, 1987
- [4] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, MIT Press, 1965.
- [5] D. Slepian and H. O. Pollack, "Prolate Spheroidal Waveforms Fourier Analysis and Uncertainty - I, in Bell Systems Technical Journal, Vol 40 pp 43-63, Jan 1961.
- [6] Ijima, Basic Theory on Normalisation of a pattern (in case of 1D pattern), Bulletin of Electrotechnical Labs, Vol 26, pp368:388 (in japanese), 1962.
- [7] J. L. Crowley and R. M. Stern, "Fast Computation of the Difference of Low-Pass Transform", IEEE Transactions on PAMI, PAMI 6(2), March 1984.
- [8] T. Lindeberg, "Feature Detection with Automatic Scale Selection", International Journal of Computer Vision, 30(2):pp79-116, 1998.
- [9] I.T. Young, L.J. van Vliet, "Recursive implementation of the Gaussian filter", Signal Processing, pages 139-151, 1997
- [10] W.T. Freeman, E.H. Adelson, "The Design and Use of Steerable Filters", Transactions on Pattern Analysis and Machine Intelligence, (PAMI), Vol 13, No. 9, pp 891-906, September 1991.
- [11] O. Chomat, V. Colin de Verdiere, D. Hall and J. L. Crowley, "Local Scale Selection for Gaussian Based Description Techniques", 6th European Conference on Computer Vision, Springer Verlag, Dublin, pp 117-133, June 2000.
- [12] D. Hall, V. Colin de Verdiere and J. L. Crowley, "Object Recognition using Coloured Receptive Field", 6th European Conference on Computer Vision, Springer Verlag, Dublin, pp 164-178, June 2000.
- [13] C. H. Chen and A. C. Kak, "A Robot Vision System for Recognition of 3D Objects in low-Order Polynomial time", IEEE Transactions on Systems, Man and Cybernetics, Vol 19 No. 6, pp 1135-1163, Nov/Dec 1989
- [14] B. V. Funt and G. D. Finlayson, "Color Constant Color Indexing", IEEE Transactions on PAMI, Vol 17. No. 5, pp 522-529, 1995.