Chapter 3

Painterly and Cubist-style Rendering using Image Salience

In this chapter we make observations on the manner in which artists draw and paint, and contrast this with the spatially local manner in which existing automatic AR techniques operate. To address this discrepancy we introduce the use of a globally computed perceptual salience measure to AR, which we apply to propose a novel algorithm for automatically controlling emphasis within painterly renderings generated from images[1]. We also propose a further algorithm which uses conceptually high level, salient features (such as eyes or ears) identified across set of images as a basis for producing compositions in styles reminiscent of Cubism[2]. These algorithms support our claim that higher level spatial analysis benefits image-space AR — specifically, enhancing aesthetic quality of output (though controlled emphasis of detail) and broadening the gamut of automated image-space AR to include artistic styles beyond stroke-based rendering.

3.1 Introduction

A drawing or painting is an artist's impression of a scene encapsulated on a two-dimensional canvas. Traditional artists often build up their renderings in layers, from coarse to fine. Coarse structure in a scene is closely reproduced, either using a trained eye, or by "squaring up"; overlaying a grid to create correspondence between image and canvas. By contrast, details within the scene are not transcribed faithfully, but are individually stylised by the artist who can direct the viewer's focus to areas of interest through judicious abstraction of the scene. Typically an artist will paint fine strokes to

[1] The salience adaptive painterly rendering technique was published in [22], and was awarded the Terry Hewitt prize for best student conference paper.

[2] An earlier version of our Cubist rendering work was published in [23].

Figure 3-1 Two examples of paintings illustrating the artist's importance-driven stylisation of detail. Non-salient texture in the background has been abstracted away, yet salient details are emphasised on the figures (left) and portrait (right) using fine strokes.

emphasise detail deemed to be important (salient) and will abstract the remaining detail away. By suppressing non-salient fine detail, yet implying its presence with coarser strokes or washes, the artist allows the viewer to share in the experience of interpreting the scene (see, for example, the paintings in Figure 3-1). Tonal variation may also be used to influence the focus, or centre of interest, within a piece [69].

The omission of extraneous, unimportant detail has been used to improve the clarity of figures in medical texts, such as Gray's anatomy [60] which consists primarily of sketches. Such sketches remain common practice in technical drawing, and were also used heavily by naturalists in the 19th century. In cinematography too, camera focus is often used to blur regions of the scene, and so redirect the viewer's attention. Both adults and children can be observed to create quick sketches and drawings of a scene by jotting down the salient lines; effectively leaving the viewer the task of interpreting the remainder of the scene. Picasso is known to have commented on what he regarded as the pain-staking nature of Matisse's art [64]; suggesting that Matisse worked by tracing the lines of a subject, then tracing the lines of the resulting drawing, and so on, each time stripping down the figure further toward its essence — "... He is convinced that the last, the most stripped down, is the best, the purest, the definitive one". In effect Matisse is iteratively refining his sketches to the few lines and strokes he deems to be salient.

We have observed (Section 2.6) that the heuristics of fully automatic AR techniques modulate the visual attributes of strokes to preserve all fine detail present in the source image. Specifically the emphasis, through level of stroke detail, afforded to a region within an artistic rendering is strongly correlated with the magnitude of high frequency

content local to that region. Such behaviour can be shown to differ from artistic practice in the general case. First, consider that fine detail, such as a fine background texture, is often less salient than larger scale objects in the foreground. An example might be a signpost set against a textured background of leaves on a tree, or a flagstone set in a gravel driveway. Second, consider that artifacts of similar scale may hold differing levels of importance for the viewer. We refer the reader back to the example of the portrait against a striped background (Figure 3-3, middle-left), in which the face and background are of similar scale but of greatly differing importance in the scene. An artist would abstract away high frequency non-salient texture, say, on a background, but retain salient detail of similar frequency characteristic in a portrait's facial features. This behaviour is not possible with automatic AR algorithms which seek to conserve all fine detail in a rendering, irrespective of its importance in the scene.

The level of detail rendered by existing automatic AR techniques is determined by the user. Constants of proportionality must be manually set which relate stroke size (and so, detail in the final rendering) with high frequency magnitude. The values of these parameters are constant over the entire image and, in practice, setting these values is a tricky, iterative process, which often requires several runs of the algorithm to produce an aesthetically acceptable rendering [58, 71, 103]. Keeping too little of the high frequency information causes salient detail to be lost, and the painting to appear blurry; keeping too much causes retention of non-salient texture (too many details in the "wrong places") which cause the output to tend toward photorealism. Moreover, if salient detail is of lower frequency magnitude than non-salient detail, then there is no acceptable solution obtainable by varying these constants — either some non-salient detail will be erroneously emphasised to keep the salient detail sharp, or some salient detail will be abstracted away in an effort to prevent emphasis of non-salient detail. This, of course, points to a flaw in the original premise; that all fine detail is salient. We thus observe that although for a given image the set of salient artifacts may intersect the set of fine scale artifacts, there may remain many salient artifacts that are not fine scale, and many fine scale artifacts that are not salient. We conclude that many images exist for which current AR methods do not emphasise some or all of the salient elements in the scene. The behaviour of current AR is at odds with that of the artist, and it is arguably this discrepancy that contributes to the undesirable impression that AR synthesised renderings are of machine, rather than true human origin.

In our opening paragraphs (Chapter 1), we argued that the notion of importance, or salience, is a relative term. When one speaks of the salience of regions in an image, one speaks of the perceptual importance of those regions *relative to that image as a whole*. Global analysis is therefore a prerequisite to salience determination; the independent

examination of local pixel neighbourhoods can give no real indication of salience in an image. The aesthetic quality of output synthesised by automatic AR would benefit from higher level (global) spatial analysis of images to drive the decision processes governing emphasis during rendering.

A further observation in Chapter 2 notes the ubiquitous trend in AR to process images at the conceptually low-level of the stroke (stroke based rendering). There are certain advantages to processing artwork at this low-level; algorithms are not only fast and simple to implement, but very little modelling of the image content is required — we have already highlighted the frequency based decision model used to guide rendering heuristics. The simplicity of this modelling admits a large potential range of images for processing. However this low-level, stroke based approach to rendering also restricts current AR to the synthesis of traditional artistic stroke based styles (such as hatching [135], stippling [43] or painterly impressionism [103]). We argue that a higher level of abstraction is necessary to extend the gamut of automated AR to encompass compositional forms of art, including abstract artistic styles such as Cubism. The successful production of such compositions is again predicated upon the development of a global image salience measure, which may be applied to identify high level salient features (for example, eyes or ears in a portrait). Such features may then be used as a novel alternative to the stroke as the atomic element in artistic renderings. In this case, the model we choose must be sufficiently general to envelope a large range of input imagery, but sufficiently high level to allow the extraction of these conceptually higher level salient features.

In the next Section, we propose a rarity based measure of salience which performs a global statistical analysis of the image to determine the relative importance of pixels. We apply this measure to develop two novel algorithms, each respectively addressing one of the two deficiencies in AR identified in the preceding paragraphs:

1. Limited ability to control emphasis, through level of detail, in renderings.

2. Limited diversity of style.

First, we propose a novel single-pass painting algorithm which paints to conserve salient detail and abstract away non-salient detail in the final rendering. Arguably this approach is more in line with traditional artistic practice, and we demonstrate the improved results (with respect to level of emphasis in the rendering) of our salience based rendering in comparison to existing AR. This algorithm serves as a pilot for salience driven painterly rendering, which we build upon to propose a salience-adaptive, relaxation based painting technique in Chapter 4, and extend to process video footage into painterly animations in Chapter 8. Second, we propose a novel rendering algorithm

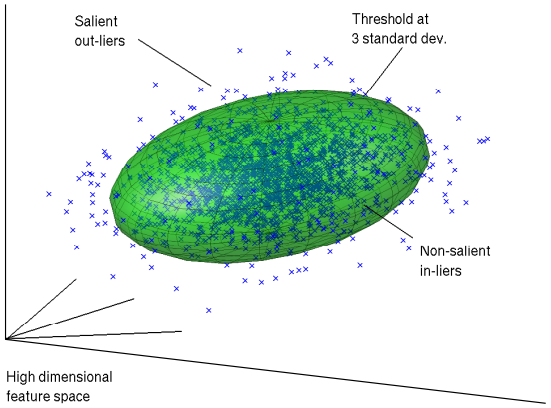

Figure 3-2 Our rarity based salience measure computes a series of derivatives for each image location, which form feature vectors in a high dimensional space. By fitting an eigenmodel to this distribution and isolating the outliers (the green hyper-ellipse boundary indicates a threshold), we may identify salient image artifacts.

capable of producing compositions in a Cubist-style, using salient features (such as eyes, ears etc.) identified across an image set. Control of the AR process is specified at the compositional, rather than the stroke based level; a further novel contribution to AR. We incorporate both these algorithms into a single system, where we apply our salience based painting algorithm to the output of the Cubist rendering algorithm to give our compositions a painterly appearance. Furthermore, we show how preferential rendering with respect to salience can emphasise detail in important areas of the composition (for example, to bring out the eyes in a portrait using tonal variation). This salience adaptation is a novel contribution to automatic image-space AR that could not be achieved without a spatially higher level, global analysis of the source image.

3.2 A Global Measure of Image Salience

We locate salient features within a single image by modifying a technique due to Walker *et al* [163], who observe that salient pixels are uncommon in an image. The basic technique is to model the statistical distribution of a set of measures associated with each pixel, and to isolate the outliers of this distribution. The pixels corresponding to these outliers are regarded as salient (Figure 3-2).

To compute these measures, $\underline{x}$, over each pixel we convolve each RGB channel of the image with a set of origin-centred 2D Gaussian derivative filters. Specifically we use 5 first and second order directional derivatives: $\partial G(x,y;\sigma)/\partial x$, $\partial G(x,y;\sigma)/\partial y$, $\partial^2 G(x,y;\sigma)/\partial x^2$, $\partial^2 G(x,y;\sigma)/\partial y^2$, and $\partial^2 G(x,y;\sigma)/\partial x \partial y$. These filters smooth the image before computing the derivative; they respond well to edge and other signals of

characteristic scale σ, but as Figure 3-3 shows, our method is more general than edge detection. We filter using octave intervals of σ, as such intervals contain approximately equal spectral power. In our implementation we use σ values of 1, 2, 4 and 8; thus with each pixel we associate a vector $\underline{x}$ of $20 = 5 \times 4 \times 3$ components.

For an image of M pixels we will have M vectors $\underline{x} \in \Re^n$, where for us $n = 60$. We assume these points are Gaussian distributed, which we represent using an eigenmodel; a simple and convenient model that works acceptably well in practice. The eigenmodel provides a sample mean $\underline{\mu}$; a set of eigenvectors each a column in orthonormal matrix $\underline{\underline{U}}$; each eigenvector has a corresponding eigenvalue along the diagonal of $\underline{\underline{\Lambda}}$. An eigenmodel allows us to compute the squared Mahalanobis distance of any point $\underline{x} \in \Re^n$:

$$d^2(\underline{x}) = (\underline{x} - \underline{\mu})^T \underline{\underline{U\Lambda U}}^T (\underline{x} - \underline{\mu}) \tag{3.1}$$

The Mahalanobis distance measures the distance between a point and the sample mean, and does so using the standard deviation (in the direction $\underline{x} - \underline{\mu}$). This provides a convenient way of deciding which sample points are salient; we use a threshold, $d^2(\underline{x}) > 9$, since 97% of normally distributed points are known to fall within 3 standard deviations of the mean. This threshold has also been shown empirically to produce reasonable results (Figure 3-3). Notice that because we look for statistical outliers we can record pixels in flat regions as being salient, if such regions are rare; a more general method than using high frequency magnitude.

Figure 3-3 demonstrates some of the results obtained when applying our salience measure to examples of real and synthetic data, and compares these results to "ground truth" importance (salience) maps which are representative of those generated by independent human observers. The global measure can be seen to out-perform local high-frequency detectors in the task of "picking out" salient artifacts, even against a textured background. We use the Sobel edge detector for comparison, as it is used by the majority of image-space AR techniques. Our measure is also shown to be sensitive to chromatic variations, whereas the Sobel edge detector used in existing AR techniques is concerned only with luminance. We observe that our global salience measure produces salience maps qualitatively closer to the ground truth that the local (Sobel) edge measure.

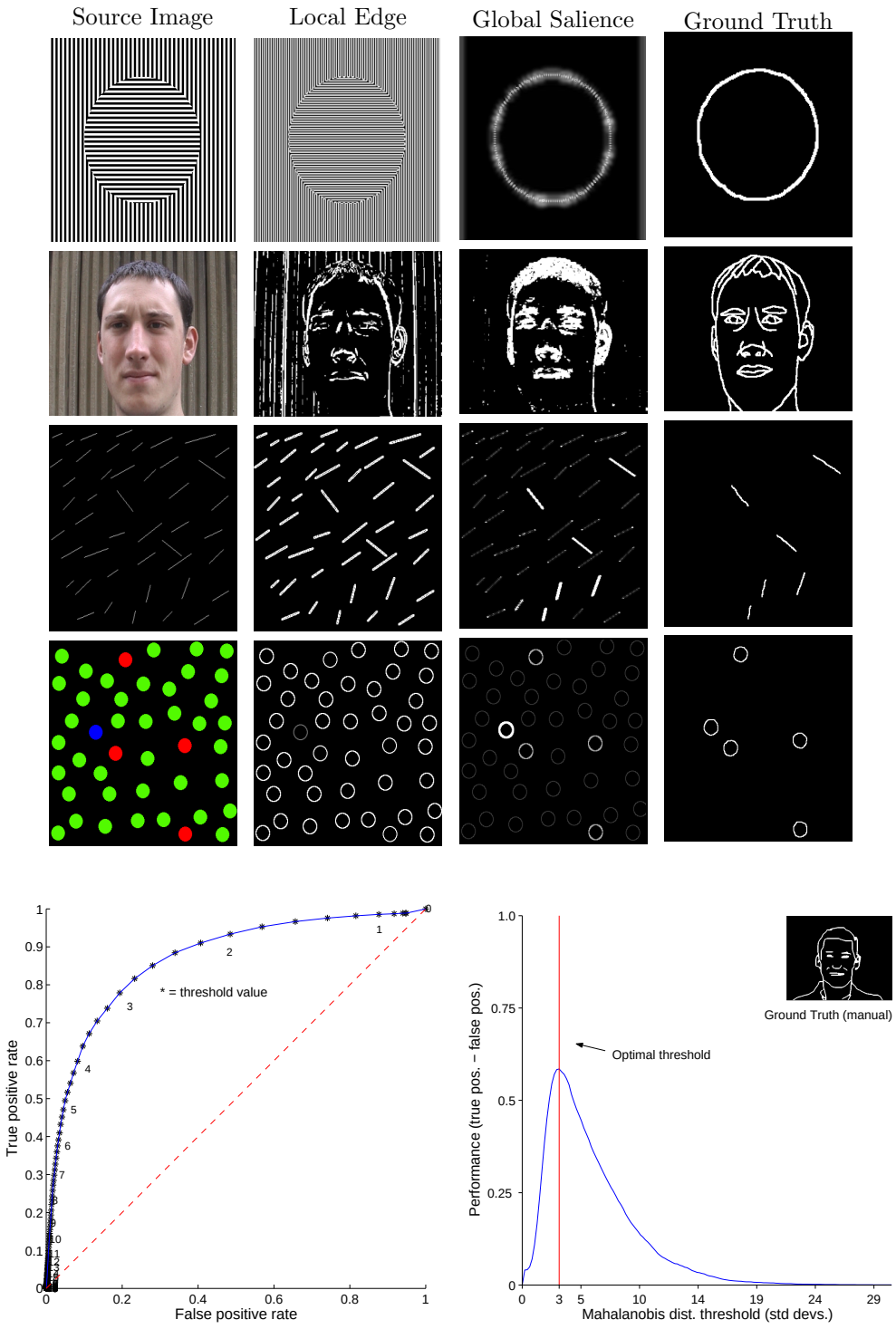

Figure 3-3 Top: Examples of real and synthetic images, processed by a local edge filter (Sobel) and our global, rarity based salience map. We observe that our approach produces maps qualitatively closer to a manually specified ground truth for image salience; we can "pick out" the circle and face where edge detection fails. Chromatic variations, scale and orientation are encapsulated in the rarity measure. Bottom: ROC curve representing sensitivity (true positive rate) vs. specificity (one minus false positive rate), as the Mahalanobis distance threshold is varied. The source image for these results is given in Figure 3-11 (middle-left). The pure chance response is plotted in dotted red. Right: Performance of the measure with various thresholds, derived from the ROC. The manually specified ground truth segmentation for this comparison is inset.

3.3 Painterly Rendering using Image Salience

We now describe a novel single-pass painterly rendering algorithm which applies our global salience measure to generate pointillist-style painterly renderings from photographs. By automatically controlling the level of emphasis in the painting (adapting the level of detail according to the salience of the region being painted), we address the first issue — aesthetic quality of rendering — raised in Section 3.1.

Our algorithm accepts a 2D image as input, and outputs a single 2D painting generated from that image. Paintings are formed by sampling a reference image at regular spatial intervals to generate a series of three-dimensional brush strokes; inverted cones with superquadric cross-section. The superquadric class of functions can be represented by the equation:

$$\left(\frac{x}{a}\right)^{\frac{2}{\alpha}} + \left(\frac{y}{b}\right)^{\frac{2}{\alpha}} = r^{\frac{2}{\alpha}} \tag{3.2}$$

where a and b are normalised constants ($a + b = 1$; $a, b > 0$) which influence the horizontal and vertical extent of the superquadric respectively, and r is an overall scaling factor. We observe that equation 3.2 reduces to the general equation for a closed elliptic curve when $\alpha = 1$, tends toward a rectangular form as $\alpha = 0$, and toward a four-pointed star as $\alpha \to \infty$. Thus the superquadrics can express a wide variety of geometric forms, using a single parameter.

Each generated conic stroke is z-buffered and the result is projected orthogonally onto the (2D) image plane to generate the final painting (Figure 3-4). There are seven parameters to each stroke; a, b, r, α (from equation 3.2), RGB colour $\underline{j}(\underline{c})$, orientation angle θ and height h. Parameter α determines the form of the stroke, and is preset by the user. Low values (< 1) of α create cross-sections of a rectangular form, giving the image a chiselled effect, whilst higher values of α produce jagged brush styles. Strokes are shaded according to the colour $\underline{c}$ of the original image at the point of sampling. Function $\underline{j}(\underline{c})$ transforms, or "jitters", the hue component of stroke colour $\underline{c}$ by some small uniformly distributed random quantity, limited by a user defined amplitude ϵ. By increasing ϵ, impressionist results similar to those of Haeberli's interactive systems [62] can be automatically produced. Further brush styles can also be generated by texturing the base of each cone with an intensity displacement map, cut at a random position from a sheet of texture; we find that this process greatly enhances the natural, "hand-painted" look of the resulting image. The remaining five stroke parameters (a, b, r, θ, and h) are calculated by an automated process which we now describe.

Stroke height h, is set proportional to image salience at the point of sampling. Higher

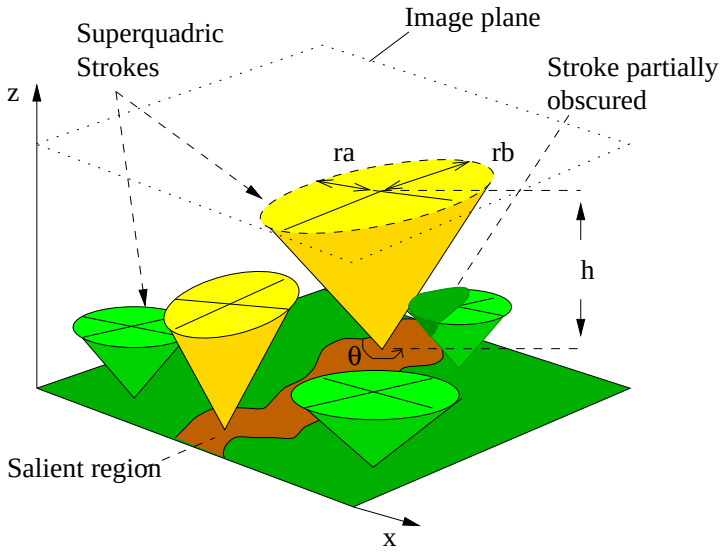

Figure 3-4 Strokes take the form of inverted cones with superquadric cross-section, and are z-buffered to produce the final painting.

salience image pixels tend to correspond to the features and detail within the image, and so produce strokes of greater height to protrude over the lower salience strokes in the z-buffer. The scale of the base of the cone, r, is set inversely proportional to salience magnitude. This causes small, definite strokes to be painted in the vicinity of artifacts corresponding to salient detail in the image. Larger strokes are used to shade non-salient areas, mimicking the behaviour of the artist. Hence our method tends to draw low, fat cones in regions of low salience, and tall, narrow cones in regions of high salience.

We also derive gradient information from the reference image, by convolving the intensity image with a Gaussian derivative of first order. Stroke orientation θ is derived from gradient orientation; the larger axis of the superquadric is aligned tangential to the edge direction. In areas where gradient magnitude is low, orientation derived in this manner becomes less reliable. We therefore vary the eccentricity of the superquadric (a, b) in relation to the magnitude of the image gradient at the position sampled. If the gradient is low, then $a \approx b$, and orientation becomes less important as the superquadric is not greatly expressed in either horizontal or vertical directions. Where image gradient is high, then $a > b$ and the superquadric stretches out. One emergent property of this approach is that strokes typically stretch along salient edges tending to merge, often causing edge highlights to appear as though produced by fewer, longer strokes. This is typical of the manner in which an artist might manually render such highlights, and adds aesthetic quality to the image.

Although we have used a global salience measure to drive emphasis in the rendering

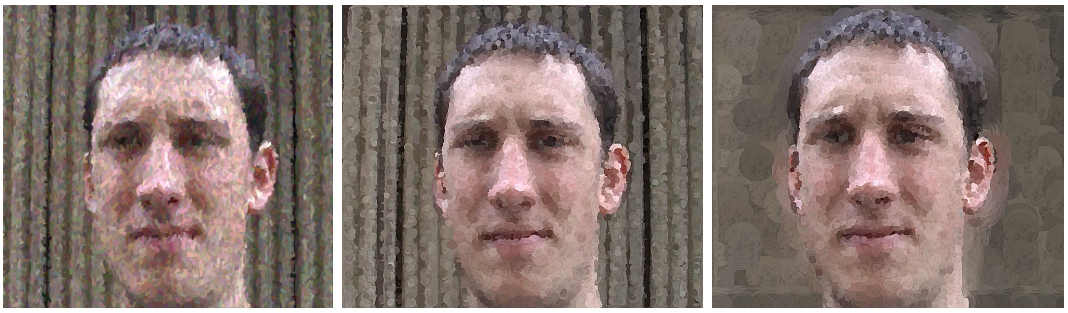

Figure 3-5 Comparison of our salience based method. Left: Results of an impressionist algorithm (due to Litwinowicz [103]). Middle: Our algorithm, but driven using Sobel response rather than global salience. Right: Our proposed salience-adaptive algorithm; non-salient background detail is abstracted away, whilst salient detail is emphasised on the face. Observe that the Sobel driven algorithms (left, middle) emphasise all high frequency detail to a similar degree, detracting from the emphasis given to the facial features. See Figure 3-3 for corresponding Sobel and salience maps.

Figure 3-6 Left: Three images of identical subject; the original image (top), painterly rendering with salience (middle), and painterly rendering without salience (bottom). The right hand column holds the salience map of the original image (top), and the edge maps of the two paintings (middle, right). Right: Strokes applied with (top) and without (bottom) salience adaptation. We make the qualitative observation that salient detail is conserved using our painterly technique.

Figure 3-7 Illustrating the application of our rendering algorithm. A section of the phone-box rendering has been magnified, demonstrating the alignment of strokes tangential to the salient window frames. Portions of the pickup truck painting with (inset, left) and without (inset, right) salience are shown. The source image and salience map are also inset below.

Figure 3-8 A "salient sketch" (right) produced by adapting our painterly technique to draw along the principal axis of each superquadric stroke — demonstrating both the close relationship between an artist's sketch and a salience map, and the potential of salience measures in driving alternative artistic rendering styles. In descending order, the original image, our automatically derived rarity based salience map, and a ground salience map, are shown on the left.

process, certain local measures have also been used to set attributes such as stroke orientation. These are inherently local properties, and we are justified in setting them as such; by contrast the concept of importance demands global analysis for computation.

3.3.1 Results and Qualitative Comparison

We present the results of applying our painting algorithm to a variety of images in Figures 3-5, 3-6 and 3-7, and demonstrate two advantages of our salience adaptive approach to painting.

Figure 3-5 contains three paintings of identical subject, painted using automatic, single-pass painterly algorithms. The right-hand painting was created using our global salience adaptive painting scheme, and demonstrates how non-salient detail (in this case, repetitive background texture) is abstracted away with coarse strokes. Salient detail has been emphasised with fine strokes, and the contrast produced against the coarser background serves to further emphasise this detail. The middle painting was generated using our algorithm, but for the purposes of illustration we have replaced our global salience measure with Sobel intensity gradient magnitude; the measure used by virtually all automatic image-space AR algorithms. Observe that the non-salient back-

ground, and salient foreground are emphasised equally, since they exhibit similar high frequency characteristics. The left-hand painting was generated using a recent daub based, painterly algorithm from the literature [103]. This algorithm is also driven by local, high frequency based heuristics and so also erroneously emphasises the non-salient background texture. A sketchy rendering of the portrait has been generated in Figure 3-8 by plotting the principal axis of each superquadric. This serves to illustrate both the alignment and placement of individual strokes, and the possibility of alternative salience driven rendering styles.

Our algorithm causes the least salient strokes to be laid down first, much as an artist might use a wash to generate wide expanses of colour in an image, and fill in the details later. Without this sensitivity to salience, the rendering procedure can obscure regions of high salience with strokes of lower salience, demonstrated by Figure 3-6. By setting the conic height h proportional to salience, salient detail is conserved within the painting — this is especially clear around the eyes and nose in Figure 3-6, left-middle. Ignoring the implicit ordering of strokes can still produce a painterly effect, but without the adaptive sensitivity to salient detail that our method provides (Figure 3-6, left-bottom). By inspection we make the qualitative observation that the majority of salient pixels in the original, and edge pixels (corresponding to detail) in the salience-painted images correspond; this is not true for the non-salience adaptive paintings.

A salience adaptive approach to painting therefore benefits aesthetic quality in two respects. Not only are salient regions painted with improved clarity (strokes from non-salient regions do not encroach upon and distort regions of greater salience — Figure 3-5), but renderings also exhibit a sense of focus around salient regions due to the abstraction of non-salient detail (Figure 3-6).

Figure 3-7 contains a gallery of paintings generated by our algorithm. The image of the pickup truck was rendered with superquadric shape parameter $\alpha = 1$. Portions of the painting rendered with and without salience adaptation are shown inset, as well as with the source image. The phone-box has been rendered with $\alpha = 0.5$; in particular we draw attention to the detail retained in the window frames (inset). Strokes have been aligned tangential to the edges of each frame, merging to create sweeping brush strokes. The strokes rendering the window glass do not encroach upon the window frames, which are more salient, and for the most-part, salient detail is conserved within the painting. A clear exception where salient detail has been lost, is within the plaque containing the words “Telephone”. Our conditioned ability to immediately recognise and read such text causes us to attribute greater salience to this region. The degradation in aesthetic quality is therefore not due to the argument for salience adaptive painting,

but rather due to the salience map of our measure diverging from the ground truth (the expectation of the viewer). This highlights the simplistic nature of our rarity driven salience measure, and suggests that one's experiences cause certain classes of artifact to be regarded as more salient than others. We return to this point later in Chapter 4, when we extend our single-pass salience adaptive painting algorithm in a number of ways — one of which is to make use of a trainable salience measure, capable of learning the classes of artifact the user typically deems to be salient.

3.4 Cubist-style Rendering from Photographs

We now describe a novel AR algorithm which addresses the second deficiency in AR identified in Section 3.1; the limited diversity of styles available by approaching AR through the low-level paradigm of stroke-based rendering. Our aim was to investigate whether aesthetically pleasing art, reminiscent of the Cubist style, could be artificially synthesised. We are influenced by artists such as Picasso and Braque, who produced art work by composing elements of a scene taken from multiple points of view. Paradoxically the Cubist style conveys a sense of motion in a scene without assuming temporal dependence between views. The problem of synthesising renderings in abstract styles such as Cubism is especially interesting to our work, since it requires a higher level of spatial analysis than currently exists in AR, in order to identify the salient features used to form stylised compositions. By *salient feature* we refer to an image region containing an object of interest, such as an eye or nose; a composition made from elements of low salience would tend to be uninteresting. We considered the following specific questions:

- How is salience to be defined so that it operates over a wide class of input images?
- How should salient features be selected from amongst many images, and how should the selected features be composed into a single image?
- How should the angular geometry common in Cubist art be reproduced?
- How should the final composition be rendered to produce a painted appearance?

Resolution of the first two questions provides the basic mechanism by which a Cubist-like image can be formed; resolution of latter two questions enhances aesthetic quality.

Our algorithm accepts one or more 2D images taken from different viewpoints as input (see Figure 3-17a), and produces a single 2D image rendered in the Cubist style. Salient artifacts within each image are first identified using our global salience measure (Section 3.2). This can produce disconnected features, which requires correction; in our case by minimal user interaction. These features are geometrically distorted. A subset

is then selected and composited, ensuring that non-selected features do not inadvertently appear in the final composition – naïve composition allows this to happen. The composition process is stochastic, and a new image may be produced on each new run of the method. An element of control may also be exerted over the composition process, affecting the balance and distribution of salient features in the final painting. The ability to influence rendering at a compositional level, rather than setting parameters of individual strokes, is a novel contribution of our method. In rendering a painting from the final composition we make use of our previously described salience based painting algorithm (Section 3.3) which treats brush strokes in a novel way, ensuring that salient features are not obscured.

We begin by registering all source images so that objects of interest, such as faces, fall upon one another; this assists the composition process in subsection 3.4.3. We threshold upon colour to partition foreground from background, and translate images so that first moments of foreground are coincident. Finally we clip the images to a uniform width and height. This step creates spatial correspondence between source images on a one-to-one basis: pixels at the same location $(x, y)^T$ in any image correspond. The remaining algorithm stages are of greater interest, and we describe each of them in turn in the following subsections.

3.4.1 Identification of Salient Features

We wish to find a set of salient features amongst the registered images. These images should be unrestricted in terms of their subject (for example, a face or guitar). In addition, we want our salient features to be relatively "high level", that is they correspond to recognisable objects, such as noses or eyes. This implies we need a definition of salience that is both general and powerful; such a definition does not currently exist in the computer vision literature, or elsewhere. However, we can make progress by choosing a definition of salience that is sufficiently general for our needs, and allow user interaction to provide power where it is needed.

We begin by applying our global salience measure to the set of source images (Section 3.2). In practice salient pixels within these images form spatially coherent clusters, which tend to be associated with interesting objects in the image, including conceptually high level features such as eyes (Figure 3-3). However, our method is general purpose, and therefore has no specific model of eyes, or indeed of any other high level feature. It is therefore not surprising that what a human regards as a salient feature may be represented by a set of disconnected salient clusters.

Given that the general segmentation problem, including perceptual grouping, remains

unsolved we have two choices: either to specialise the detection of salient regions to specific classes of source images, such as faces (see the fully automatic case study described later in Section 3.5); or to allow the user to group the clusters into features. We adopt the latter approach for its power and simplicity: powerful because we rely on human vision, and simple not only to implement but also to use. We allow the user to draw loose bounding contours on the image to interactively group clusters. This mode of interaction is much simpler for the user than having to identify salient features from images *ab initio*; that is with no computer assistance. Feature grouping is also likely to be consistent between source images, because our salience measure provides an objective foundation to the grouping. The contour specified by the user forms the initial location for an active contour (or "snake"), which is then iteratively relaxed to fit around the group of salient clusters. Active contours are parametric splines characterised by an energy function E_{snake}; the sum of internal and external forces [91]. Internal forces are determined by the shape of the contour at a particular instant, and external forces are determined by the image upon which the contour lies. Here the spline is defined by a parametric function $\underline{v}(s)$:

$$E_{snake} = \int_0^1 E_{snake}(\underline{v}(s))ds \tag{3.3}$$

$$E_{snake} = \int_0^1 E_{internal}(\underline{v}(s)) + E_{external}(\underline{v}(s))ds \tag{3.4}$$

During relaxation we seek to minimise the contour's energy by iteratively adjusting the position of the spline control points, and thus tend toward an optimal contour fit to the salient feature. In our case energy is defined as:

$$E_{internal} = \alpha\left|\frac{d\underline{v}}{ds}\right|^2 + \beta\left|\frac{d^2\underline{v}}{ds^2}\right|^2 \tag{3.5}$$

$$E_{external} = \gamma f(\underline{v}(s)) \tag{3.6}$$

where the two terms of the internal energy constrain the spacing of control points and curvature of the spline respectively. The external energy function is simply the sum of salience map pixels bounded by the spline $\underline{v}(s)$, normalised by the number of those pixels. Constants α, β and γ weight the importance of the internal and external constraints and have been determined empirically to be 0.5, 0.25 and 1. In our application we fit an interpolating (Catmull-Rom) spline [51] through control points to form the parametric contour $\underline{v}(s)$. We assume that the user has drawn a contour of approximate correct shape around the feature clusters; the weighting constants have been chosen

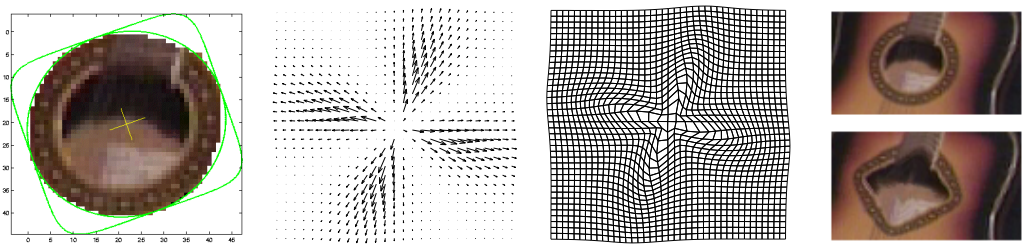

Figure 3-9 Four stages of a geometric warp where $\alpha' = 0.3$. From left to right: (a) the source and target superquadrics, fitted about a salient feature; (b) the continuous forward vector field; (c) the mesh of quadrilaterals (mapped pixel areas); (d) the final distorted image.

to promote retention of initial shape in the final contour. Relaxation of the snake proceeds via an algorithm adapted from Williams [167], in which we blur the salience map heavily in early iterations and proportionately less on subsequent iterations. This helps prevent the snake snagging on local minima early in the relaxation process. Our initial approach [23] made use of a convex hull based grouping technique to form salient clusters, however this precluded the possibly of accurately extracting concave features. Further advantages of the snake segmentation method are a tighter, more accurate fit to features, and greater robustness to noise. There is also a lesser degree of sensitivity upon the initial positioning of the contour, since the snake shrinks to fit the exterior boundary of the salient pixel cluster.

The union of the salient features identified in each source image forms the set of salient features we require, which we call $\mathcal{F}$. In addition to grouping clusters into features, the user may also label the features. These labels partition the set of all salient features $\mathcal{F}$ into equivalence classes, such as "eyes", providing a useful degree of high level information (these classes represent a simple model of the object in the picture). We make use of $\mathcal{F}$, and associated equivalence classes, throughout the remaining three stages of our algorithm.

3.4.2 Geometric Distortion

We now wish to distort the identified features, in $\mathcal{F}$, to produce the more angular forms common in Cubist art. Our approach is to construct a continuous vector field $\mathcal{V}$ over each source image, which is a sum of the contributions made by distorting the set of all features $f \in \mathcal{F}$ belonging to that image. That is, we define a vector-valued distortion function $\underline{g} : \Re^2 \mapsto \Re^2$, so that for every point $\underline{u} \in \Re^2$, we have $\underline{g}(\underline{u}) = \underline{u} + \mathcal{V}(\underline{u})$ where

$$\mathcal{V}(\underline{u}) = \sum_{\phi \in f} \underline{d}_\phi(\underline{u}) \tag{3.7}$$

To define a particular distortion function $\underline{d}_\phi(.)$ we fit a superquadric about the perime-

ter of feature ϕ, then transform that fitted superquadric to another of differing order; thus specifying a distortion vector field $\underline{d}_\phi(\Re^2)$. We now describe the details of this process.

Recall equation 3.2 in which the superquadric class of functions may be represented in Cartesian form by:

$$\left(\frac{x}{a}\right)^{\frac{2}{\alpha}} + \left(\frac{y}{b}\right)^{\frac{2}{\alpha}} = r^{\frac{2}{\alpha}} \tag{3.8}$$

We use a parametric form of equation 3.2 determined by an angle θ about the origin, by which we correlate points on the perimeter of one superquadric with those on another.

$$x = \frac{r\cos(\theta)}{\left(|\cos(\theta)/a|^{\frac{2}{\alpha}} + |\sin(\theta)/b|^{\frac{2}{\alpha}}\right)^{\frac{\alpha}{2}}} \tag{3.9}$$

$$y = \frac{r\sin(\theta)}{\left(|\cos(\theta)/a|^{\frac{2}{\alpha}} + |\sin(\theta)/b|^{\frac{2}{\alpha}}\right)^{\frac{\alpha}{2}}} \tag{3.10}$$

We calculate the distortion for a given feature by fitting a general superquadric of order α, and warping it to a target superquadric of new order α'. Features whose forms differ from this target superquadric are therefore distorted to a greater degree than features that already approximate its shape; thus each feature boundary converges toward the geometric form specified by α'. Typically we choose $\alpha' < 1$ to accentuate curves into sharp angles. The initial superquadric is fitted about the bounding pixels of the feature using a 6-dimensional Hough transform based search technique described in Appendix A.1. We find this global fitting method suitable for our purpose due to its relatively high tolerance to noise.

Recall the distortion function $\underline{d}_\phi(.)$; we wish to produce a displacement vector $\underline{v}$ for a

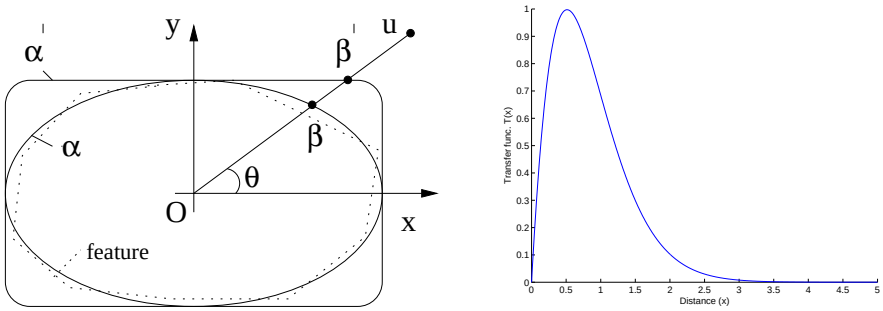

Figure 3-10 Left: The fitted and target superquadrics, described by α and α' respectively. Intersection with line $\vec{Ou}$ is calculated using angle θ. Right: The decay function (equation 3.12) used to dampen the vector field magnitude.

given point $\underline{u} = (u_x, u_y)$. We first calculate the points of intersection of line $\underline{\vec{Ou}}$ and the two superquadric curves specified by α and α', where $\underline{O}$ is the origin of both superquadrics (these origins are coincident). We derive the intersections by substituting a value for $\theta = arctan(u_y/u_x)$ into equations 3.9) and (3.10. We denote these intersection points by $\underline{\beta}$ and $\underline{\beta'}$ respectively (see Figure 3-10, left). The vector $\underline{\beta'} - \underline{\beta}$ describes the maximum distortion in direction θ. We scale this vector by passing the distance (in superquadric radii) of point $\underline{u}$ from the origin, through a non-linear transfer function $\mathcal{T}(.)$. So, for a single feature ϕ:

$$\underline{d}_\phi(\underline{u}) = \mathcal{T}\left(\frac{|\underline{u} - \underline{O}|}{|\underline{\beta} - \underline{O}|}\right)(\underline{\beta'} - \underline{\beta}) \tag{3.11}$$

The ideal characteristics of $\mathcal{T}(x)$ are a rapid approach to unity as $x \to 1$, and a slow convergence to zero as $x \to \infty$. The rise from zero at the origin to unity at the superquadric boundary maintains internal continuity, ensuring a topologically smooth mapping within the superquadric (Figure 3-10, right). Convergence to zero beyond unit radius mitigates against noticeable distortion to surrounding areas of the image that do not constitute part of the feature. The Poisson distribution function (equation 3.12) is a suitable $\mathcal{T}$, where $\Gamma(.)$ is the gamma function [123] and λ is a scaling constant.

$$\mathcal{T}(x) = \frac{\lambda^{x\mathrm{e}^\lambda}}{\Gamma(x)} \tag{3.12}$$

Recall from equation 3.7 that we sum the individual vector fields of each feature belonging to a specific source image, to construct the overall vector field for that image. With this field defined, we sample those points corresponding to the corners of every pixel in the source image, and so generate their new locations in a target image. This results in a mesh of quadrilaterals, such as that in Figure 3-9*c*. Mapping each pixel area from the original bounded quadrilateral to the target bounded quadrilateral yields the distorted image.

The distortion process is repeated for each source image, to produce a set of distorted images. At this stage we also warp the bounding polygon vertices of each feature, so that we can identify the distorted salient features $\mathcal{F}'$. For reasons of artistic preference, we may wish to exercise control to use different values of α' for each equivalence class; for example, to make eyes appear more angular, but leave ears to be rather more rounded.

We draw attention to issues relating to the implementation of our method; specifically that the feature distortion stage can be relatively expensive to compute. This bottleneck can be reduced by: (a) precomputing the transfer function $\mathcal{T}(.)$ at suitably small

discrete intervals, and interpolating between these at run-time; (b) using a fast but less accurate method of integrating distorted pixel areas such as bilinear interpolation. In both cases we observed that the spatial quantisation induced later by the painterly rendering stage mitigates against any artifacts that may result.

3.4.3 Generation of Composition

We now describe the process by which the distorted salient features are selected from $\mathcal{F}'$ and composited into a target image. Specifically we wish to produce a composition in which:

- The distribution and balance of salient features composition may be influenced by the user.
- Features do not overlap each other.
- The space between selected salient features is filled with some suitable non-salient texture.
- Non-salient regions are "broken up" adding interest to the composition, but without imposing structure that might divert the viewer's gaze from salient regions.

A subset of the distorted salient features $\mathcal{F}'$ are first selected via a stochastic process. These chosen features are then composited, and an intermediary composition produced by colouring uncomposited pixels with some suitable non-salient texture. Large, non-salient regions are then fragmented to produce the final composition.

Selection and Composition

We first describe the process by which a subset of distorted salient features in $\mathcal{F}'$ are selected. We begin by associating a scalar $s(f)$ with every feature $f \in \mathcal{F}$:

$$s(f) = A(f) \cdot T(\mathcal{E}(f)) \tag{3.13}$$

in effect the area of the feature $A(f)$ weighted by a function $T(.)$ of the fractional size of the equivalence class to which it belongs (which we write as $\mathcal{E}(f)$). By varying the transfer function $T(.)$, the user may exercise control over the balance of the composition. We use:

$$T(x) = x^{\beta} \tag{3.14}$$

Figure 3-11 Left: Three source images used to create a Cubist portrait. Right: Features selected from the set $\mathcal{F}'$ via the stochastic process of Section 3.4.3 with balance parameter $\beta = 1$. Notice that the number of facial parts has a natural balance despite our method having no specific model of faces; yet we still allow two mouths to be included. Pixels not yet composited are later coloured with some suitable non-salient texture.

β is a continuous user parameter controlling visual "balance" in the composition. If set to unity, features are distributed evenly in similar proportion to the equivalence classes of the original image set. By contrast $\beta = 0$ introduces no such bias into the system, and a setting of $\beta = -1$ will cause highly unbalanced compositions, in which rarer classes of feature are more likely to be picked.

We treat each scalar $s(f)$ as an interval, and concatenate intervals to form a range. This range is then normalised to span the unit interval. We choose a random number from a uniform distribution over $[0, 1]$, which falls in a particular interval, and hence identifies the corresponding feature. Features of larger area with large associated scalar values ($s(f)$) tend to be selected in preference to others, which is a desirable bias in our stochastic process. The selected feature is removed from further consideration, and included in a set $\mathcal{C}$, which is initially empty.

This selected feature may intersect features in other images, by which we mean at a least one pixel (i, j) in the selected feature may also be in some other feature in some other image (recall the registration process aligns pixels to correspond on a one-to-one basis). Any features that intersect the selected feature are also removed from further consideration, but are not placed in the set $\mathcal{C}$.

We have found the choice of $\beta = 1$ to produce aesthetically pleasing renderings. In this case the process is biased toward producing a set $\mathcal{C}$ containing features whose equivalence classes are similar in proportion to the original source images. For example, if the original subject has two eyes and a nose, the algorithm will be biased toward

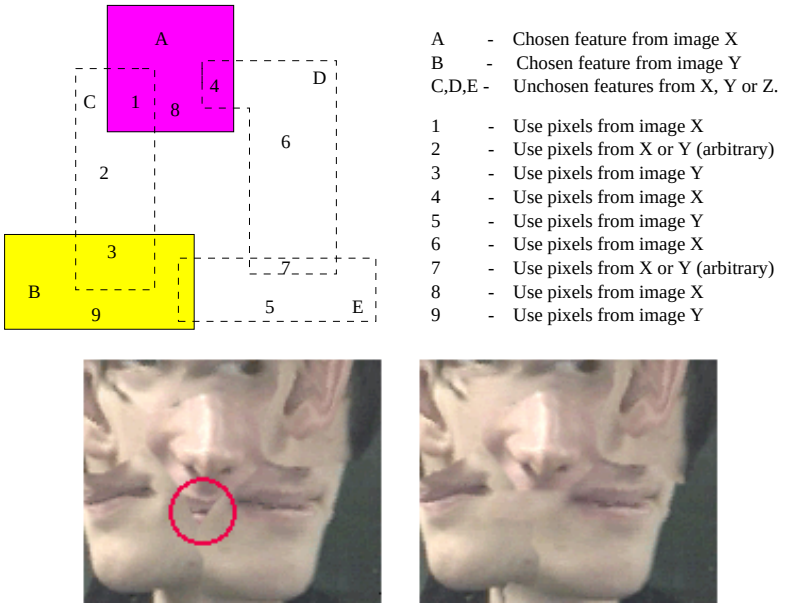

Figure 3-12 (a) potential intersections between features (top); (b) final compositions without (left) and with (right) the second stage of processing.

producing a composition also containing two eyes and a nose, but deviation is possible, see Figure 3-11.

The second step of our process is concerned with the composition of the chosen features in $\mathcal{C}$ to produce the final image. We begin this step by copying all chosen features into a target plane, producing a result such as Figure 3-11. In order to complete the composition we must determine which image pixels have not yet been composited, and colour them with some suitable non-salient texture.

An initial approach might be to compute a distance transform [145] for each non-composited pixel, which determines its distance to the nearest feature. The corresponding pixel in the distorted source image containing this nearest feature is used to colour the uncomposited pixel. This produces similar results to a Voronoi diagram, except that we seed each Voronoi segment with a region rather than a point. Unfortunately this initial approach is unsatisfactory: under some circumstances regions may be partially textured by unchosen salient features, and images such as Figure 3-12*b* (left) may result. To mitigate against partial mouths and similar unappealing artifacts requires greater sophistication, which we introduce by performing a second set of intersection tests.

We copy each of the unchosen features onto the image plane, and test for intersection with each of the chosen features $\mathcal{C}$. If an unchosen feature u intersects with a cho-

Figure 3-13 Illustrating composition from a collection of warped salient features in the guitar image set. Composition balance parameter $\beta = 1$.

sen feature c, we say that *'c holds influence over u'*. Unchosen features can not hold influence over other features. By examining all features, we build a matrix detailing which features hold influence over each other. If an unchosen feature u is influenced by exactly one chosen feature c, we extend feature c to cover that influenced area. We fill this area by copying pixels from corresponding positions in the distorted image from which c originates. Where an unchosen feature is influenced by several chosen features, we arbitrarily choose one of these chosen features to extend over the unchosen one (Figure 3-12a, region 2). However, we do not encroach upon other chosen regions to do this – and it may be necessary to subdivide unchosen feature areas (Figure 3-12a, regions 1, 3 and 4). Only one case remains: when two unchosen features intersect, which are influenced by features from two or more differing source images (Figure 3-12a, region 7). In this case we arbitrarily choose between those features, and copy pixels from the corresponding distorted source image in the manner discussed.

We now perform the previously described distance transform procedure on those pixels not yet assigned, to produce our abstract composition.

Fragmentation of Non-salient Areas

The composition produced at this stage (Figure 3-14a, left) is often composed of pieces larger than those typically found in the Cubist paintings. We wish to further segment non-salient regions to visually "break up" uninteresting parts of the image, whilst avoiding the imposition of a structure upon those areas.

We initially form a binary mask of each non-salient segment using information from the previous distance transform stage of Section 3.4.3, and calculate the area of each segment. We then average the area of the chosen salient features $\mathcal{C}$, to produce a desired "segment size" for the composition. Each non-salient segment is fragmented into n pieces, where n is the integer rounded ratio of that segment's area to the desired segment size of the composition. To perform the segmentation we produce a dense point cloud of random samples within the binary mask of each non-salient segment. Expectation maximisation [41] is used to fit n Gaussians to this point cloud. We then

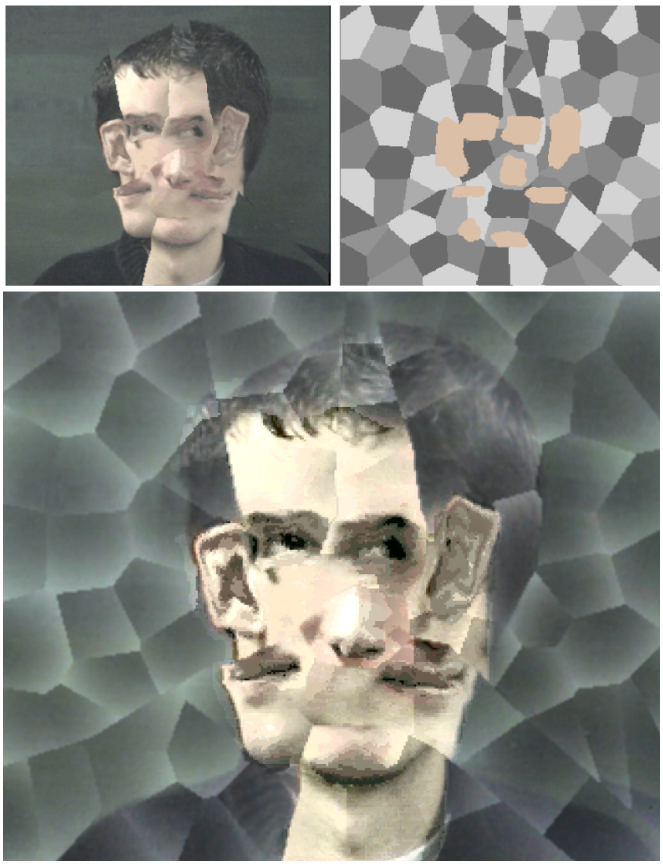

Figure 3-14 (a) Composition after application of steps in Section 3.4.3 exhibiting large non-salient segments (left) and a uniquely coloured finer segmentation (right) (b) Results of finer segmentation and shading of non-salient areas in the composition.

calculate the Gaussian centre to which each pixel within a given mask is closest; a Voronoi diagram is thereby constructed, the boundaries of which subdivide the non-salient segment being processed into multiple non-salient *fragments*.

Each of the non-salient fragments must now be shaded to break up the composition. We choose a point, or "epicentre" along each fragment's boundary, and decrease the luminosity of pixels within that fragment proportional to their distance from the epicentre (see Figure 3-15). The result is a modified intensity gradient across each fragment, simulating light cast over a fragment's surface. In practice it is desirable that no two adjacent fragments have an intensity gradient of similar direction imposed upon them; doing so induces a noticeable regular structure in non-salient areas, which can divert the viewer's attention from the more interesting *salient* features elsewhere in the composition. Placement of the epicentre at a random location upon the boundary produces too broad a range of possible gradient directions, causing shading to appear as noise. We therefore restrict shading to a minimal set of directions, calculated in the following manner.

A region adjacency graph is constructed over the entire composition; each non-salient fragment corresponds to a node in the graph with vertices connecting segments adjacent in the composition. We then assign a code or "colour" to each node in the graph, such that two directly connected nodes do not share the same colour. Graph colouring is well-studied problem in computer science, and an minimal colour solution is known to be NP-hard to compute. We therefore use a heuristic based approximation which is guaranteed to return a colouring in P-time, but which may not be minimal in the number of colours used (see Appendix A.2 for details). The result is that each fragment is assigned an integer coding in the interval $[1, t]$, where t is the total number of colours used by our approximating algorithm to encode the graph.

The result of one such colouring is visualised in Figure 3-14*a*. The epicentre of each fragment is placed at the intersection of the fragment's boundary and a ray projected from the centroid of the fragment at angle θ from vertical (Figure 3-15), where θ is determined by:

$$\theta = 2\pi \left(\frac{segment\ code}{t} \right) \tag{3.15}$$

This expression guarantees placement of the epicentre at one of t finite radial positions about the boundary of the segment, as the segment coding is an integer value.

The introduction of additional segmentation, and therefore edge artifacts, into non-salient areas of the composition can have the undesired effect of diverting a viewer's gaze from salient features present in the picture. We mitigate against this effect in two ways. First, we convolve the non-salient regions of the image with a low-pass filter kernel such as a Gaussian. This has the effect of smoothing sharp edges between fragments, but conserving the more gradual intensity differential over each non-salient fragment's

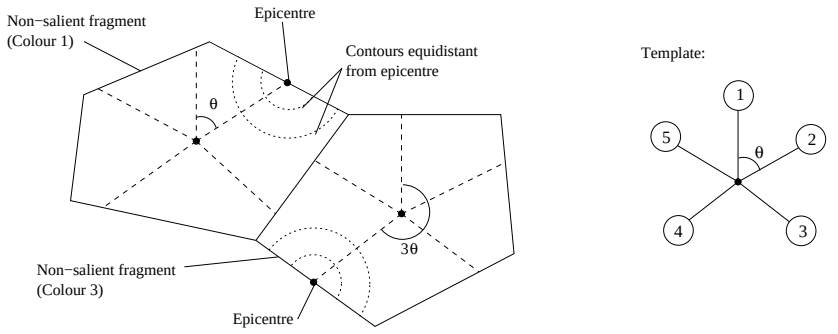

Figure 3-15 The geometry of two adjacent non-salient fragments, and a single template determining the location of the epicentre within a fragment. The epicentre of a fragment of colour i lies at the intersection of that fragment's boundary and the ith template spoke.

Figure 3-16 Comparison of two painted compositions with (left) and without (right) preferential shading treatment to salient regions. Observe how salient features such as the eyes are brought out within the salience adaptive composition; tonal variation has been used to emphasise the salient regions via histogram equalisation.

surface. This also proves advantageous in that the jagged edges of lines partitioning fragments are smoothed. Second, we use a variation upon histogram equalisation [145] to boost contrast within the foreground of the composition (determined during image registration), causing features such as eyes or noses to "stand out" from the softened segmentation boundaries. Specifically, we calculate the transfer function between the luminosities of the source and equalised compositions. For each pixel in the composition we then interpolate between these luminosities proportional to that pixel's salience; thus contrast is boosted in more salient areas of the composition, greatly improving the aesthetics of the painting (Figure 3-16).

This produces a final composition such as that of Figure 3-14*b*. We draw attention to the fact that major segmentation lines (produced by the steps of Section 3.4.3) and salient features remain unaffected by this final segmentation of the composition.

3.4.4 Applying a Painterly Finish

The final stage of our algorithm is concerned with creating a painterly effect on the generated composition, to which there are two sub-stages: colour quantising, and brush stroke generation.

The colour quantising step should be performed prior to composition, but is described here for the sake of clarity. We use variance minimisation quantisation [174], to reduce the colour depth of three independent areas within the image: the distorted salient features ($\mathcal{F}'$); the foreground of each distorted image; and the background of each distorted image. Distinction between foreground and background is made by thresholding upon a simple characteristic property of the image, such as hue or intensity (as was performed during image registration). Our motivation to quantise follows the observation

that an artist typically paints with a restricted palette, and often approximates colours as a feature of the Cubist style [81]. We allow a level of control over this effect by differentiating the level of quantisation over the various image components, and have found that heavy quantisation of the features and foreground, contrasted by a lesser degree of background quantisation can produce aesthetically pleasing effects.

At this stage we optionally introduce false colour to the image. Artists such as Braque and Gris often painted in this manner, contrasting shades of brown or grey with yellows or blues to pick out image highlights. We use a look-up table based upon a transfer function, which generates a hue and saturation for a given intensity, calculated from the original input colour. Typically we define this function by specifying several hue and saturation values at various intensities, and interpolate between these values to produce a spectrum of false colour to populate the look-up table.

The second step of the rendering process concerns the generation of "painted" brush strokes, using our previously described salience adaptive painterly technique (Section 3.3). This ensures that strokes from non-salient regions to do encroach upon salient features during this final rendering step.

3.4.5 Results of Cubist Rendering

We present the results of applying our Cubist algorithm to three image sets; a portrait, a guitar, and a nude. These subjects were popular choices for artists of the Cubist period, and we use them to demonstrate the processes of composition, distortion, and painting respectively.

The original source image sets were captured using a digital video camera, and are given in Figure 3-17a. Figure 3-17b presents the results of processing the portrait images; salient features were the ears, eyes, nose and mouth. Figures 3-17b_1 and 3-17b_2 were created by successive runs of the algorithm, using identical distortion parameters; the stochastic nature of feature selection produces varying compositions in the same visual style. Figure 3-17b_3 demonstrates the consequence of relaxing the constraints which maintain proportion between equivalence classes during composition; equivalence classes are no longer proportionally represented; in this case parameter $\beta = 0$.

The nude has been rendered with minimal distortion; salient features were the eyes, nose, mouth, chest and arm. False colour has been introduced to Figure 3-17c_2, using the complementary colours of blue and orange to contrast highlight and shadow. Many abstract artists make use of complementary colour pairs in a similar manner.

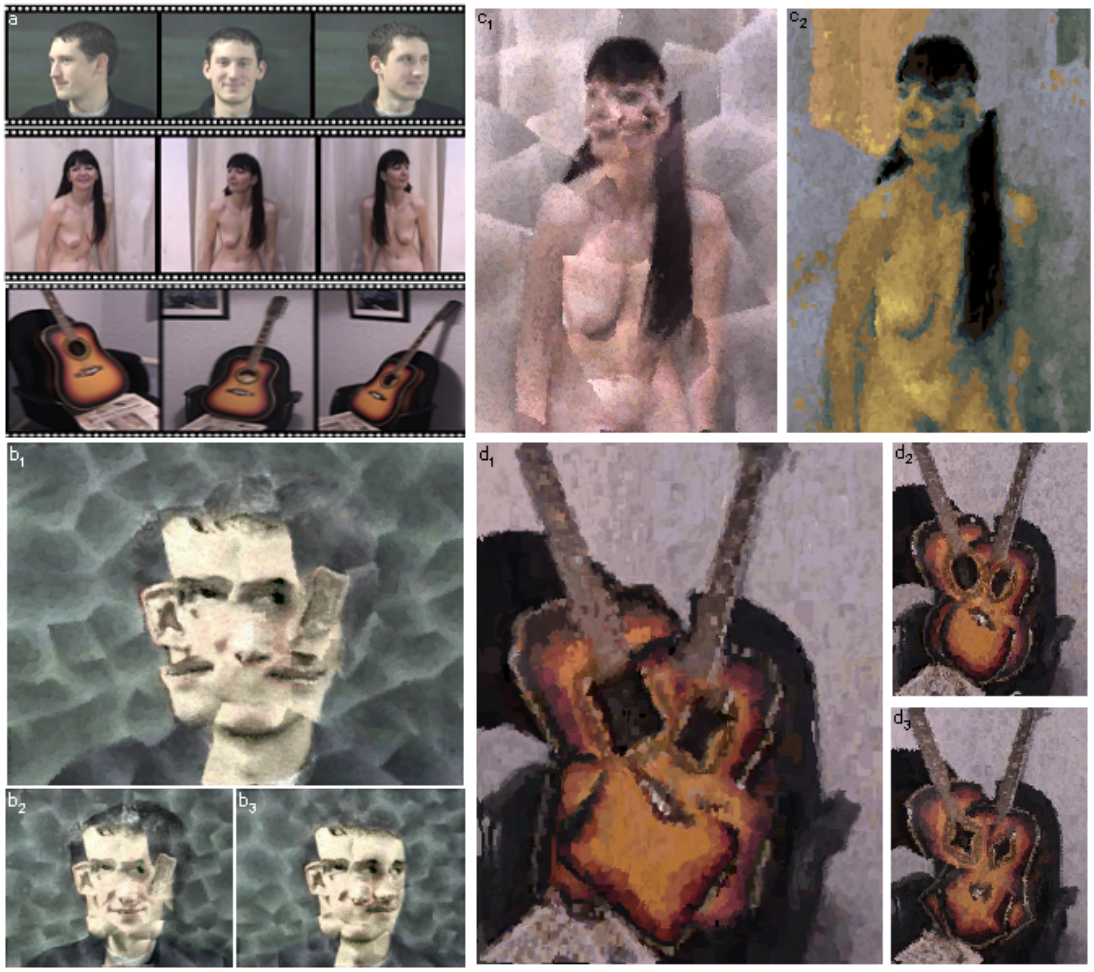

Figure 3-17 A gallery of images illustrating the application of our rendering algorithm.

Paintings produced from the guitar images are presented in Figure 3-17*d*; salient features were the hole, neck, bridge and chair arms. Figures 3-17d_1, 3-17d_2, and 3-17d_3 are identical compositions rendered with different distortion and painting parameters. The values of distortion parameter α' for each of the renderings is 0.5, 1, and 2 respectively. Notice how the hole in the guitar changes shape, from rectangular to star-like. By changing only these parameters, a varied range of styles are produced. The finer segmentation of non-salient regions was not performed on these renderings, to allow clear demonstration of distortion effects.

New year honours
James Lovelock is among academe's acclaimed 9

Briefing notes
What should the government address in its strategy paper? 6

Rising stars
Which big name is destined to find mainstream fame? 18

THE TIMES
HIGHER
EDUCATION SUPPLEMENT

www.thes.co.uk
JANUARY 3 2003 No.1,570 £1.40

Surrey seeks escape from state control

Caroline Davis

Surrey University could become Britain's first public university to opt out of government control and rely on independent funding.

In 2000-01, the university was among the institutions that received the lowest proportion of their income from the Higher Education Funding Council for England.

New figures released by Hefce show that, on average, institutions received 41 per cent of their income from the funding council. Only 17 universities now derive the majority of their funding from the council. Just 25 per cent of Surrey's income came from Hefce, down from 28 per cent three years earlier.

Vice-chancellor Patrick Dowling said that the university listed government higher education policy as one of the highest risk factors in its strategy. "Something has to be done. It could reach the stage where we feel strongly enough [to become independent]. I would like to be able to keep our options open," he said.

He told *The THES*: "It is not easy going private overnight. But if we were given a major sum or endowment, we would be one of the first ones to say let's have a go."

Until recently, Surrey's undergraduate degree in dance was privately funded by student fees and was never short of applicants. But Professor Dowling said tuition fees should not rise as this could deter students.

He said the university could operate totally independently from government, being research-led with high-quality teaching at both undergraduate and postgraduate levels.

Since suffering huge budget cuts in the early 1980s, Surrey has worked to distance itself from government funding. It increased the proportion of postgraduate and overseas students (who pay fees) and set up a "land bank", profiting from sales, most recently for the building of a hospital. Surrey was also the first UK university to own a science research park, from which it nets about £4 million a year in rent.

In 2000-01, the university received £29 million of its £116 million income from Hefce grants. The rest came from academic fees and support grants, research grants and contracts, endowments, rental income from its research park, and its ten subsidiary companies.

Other institutions, however, said they would be unable to turn down millions of pounds in funding council support.

Imperial College London received 29 per cent of its income from Hefce last year. But despite raising the prospect of tuition fees of up to £15,000, the college said it could not afford to give up government support.

Rodney Eastwood, Imperial's director of planning and information, said: "We could not be independent from the £60 million Hefce research grant." He said that including grants from the research councils, Imperial was 50 per cent funded by the public.

Cambridge University derived 32 per cent of its income — £140 million — from Hefce. Treasurer Joanna Womack said that the endowment required to generate a similar income was unrealistic.

"At 4 per cent, that represents a capital sum of £3.5 billion," she said. "We could not raise that, and even if we could, such a sum would just leave us where we are at present, whereas, like all universities, we need to grow our income in order to cover increased costs. What we really need is increased Hefce funding for both research and teaching, guaranteed for several years so that we can make sensible plans."

Michael Sterling, vice-chancellor of Birmingham University, suggested that he would like to move towards more independent funding. With one-third of its income from Hefce, he said there was no possibility of the university opting out. "That's still too much on our

continued on page 2

Universities that receive least and most Hefce funding as percentage of income

Institution	00-01	97-98	Change
London Business School	6	9	-33%
LSE	19	23	-17%
City	24	27	-11%
Surrey	25	28	-11%
Cranfield	16	17	-6%
North London	57	58	-2%
London Guildhall	60	61	-2%
Manchester Met	59	58	2%
Ulster	58	55	5%
Lincoln	61	55	11%

Source: Hefce

Time to work on the image, Mr Clarke?

Steve Farrar

Culture minister and self-appointed art critic Kim Howells will have to curb his tongue this time. Just weeks before his colleague Charles Clarke makes clear his vision for the future of higher education, the education secretary's features have been distorted in the cubist style (pictured) to demonstrate new computer technology at Bath University.

The technique, developed by postgraduate student John Collomosse and lecturer Peter Hall, involves minimal human intervention to turn photographic images into the sort of artworks once created by Picasso.

The researchers used a selection of photographs of Mr Clarke, taken from different viewpoints and supplied by *The THES*.

Dr Hall said: "In order to draw, you have to be able to see." The software, he explained, picks out individual elements, which it geometrically distorts before reassembling them into a coherent composition. It then turns a selection of dots into brush strokes that work around important aspects of the picture, giving a painterly effect.

Professional artists have judged the project's output to be of a high aesthetic quality, and the Bath team has entered work in computer art competitions.

The research was supported by the Engineering and Physical Sciences Research Council and will be published in the journal *Transactions*.

***THES* diary, page 13**

Figure 3-18 A Cubist-style image of Charles Clark MP appeared in the Times Higher Educational Supplement [3rd January 2003], and was rendered from a set of photographs supplied by Times reporter Steve Farrar (below).

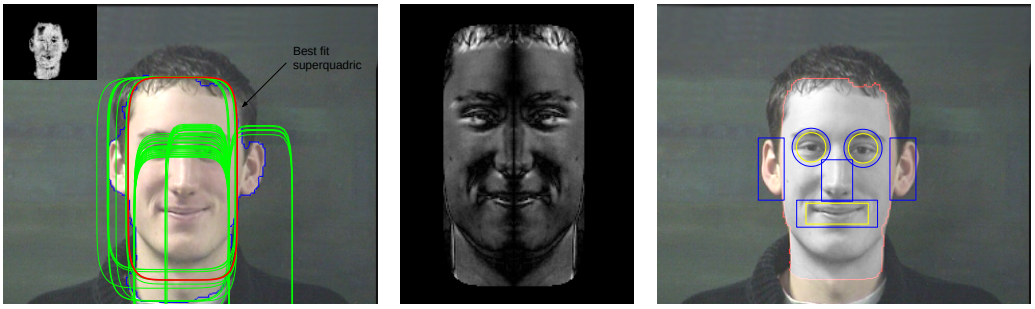

Figure 3-19 Illustrating our ad-hoc facial registration technique. Left: Images are processed using a hue/saturation based eigen-model to identify flesh coloured areas (inset). Candidate faces (green) are found using our Hough based search strategy (Appendix A.1), which operates on edge pixels identified in the image (blue) following eigen-analysis. Middle: Difference map used to quantify the symmetry expressed by image data along the principal axis of a candidate superquadric. Right: A rigid template containing eyes and mouth (yellow) is fitted to the image data by translation and scaling within the ortho-normal basis of the best fit superquadric. Snake initial contour locations are also transformed with the template (blue), and subsequently relaxed to identify salient features (Section 3.4.1).

3.5 Personal Picasso: Fully Automating the Cubist Rendering System

The general segmentation problem unfortunately prohibits the fully automatic extraction of salient features from a general image. However certain classes of image are well studied in Computer Vision, and can be automatically segmented into such features. In this section we describe an implementation of our Cubist rendering system which adapts techniques from the Vision literature to locate the features of a face within a single video frame. These features are then tracked through subsequent video frames, and so substitute the interactive elements of our Cubist algorithm to create a fully automatic rendering process. The motivation for this case study is to produce a "Personal Picasso" system capable of rendering Cubist portraits from video. Potential applications for this system might include installation as a feature in commercial photo booths. We describe a proof of concept implementation of such a system in the following subsections.

3.5.1 An Algorithm for Isolating Salient Facial Features

We begin by describing an ad-hoc technique for locating salient facial features within a single image. The first stage of processing involves the location of the face within the image; we assume that a full frontal image of a single face will always be present. The second stage localises the various facial features, e.g. eyes, mouth within the identified face.

Locating the face

Faces are located on the basis of their distinctive colour signature and shape, using a ad-hoc process which draws upon both previous colour-blob based face location strategies (for example, [125]) and the Hough transform based, geometric approach of [106].

It is well known that, despite natural variations in skin tone and colour, skin pigments tend to form a tight cluster in 2D hue/saturation space [49]. Prior to processing, we perform a one-time training process which fits an eigenmodel to various samples of skin colour taken from photographs; this has empirically proven to be a satisfactory model of the unimodal distribution of pigments. The eigenmodel is specified by a mean $\underline{\mu}$, eigenvectors $\underline{\underline{U}}$, and eigenvalues $\underline{\underline{\Lambda}}$. When processing a novel source image for face location, we compute the Mahalanobis distance of each novel pixel's colour with respect to the trained eigenmodel. The Mahalanobis distance $L(\underline{c})$ of a particular pixel with colour $\underline{c} = (c_h, c_s)^T$ (where c_h is the colour hue component, and c_s the saturation — both components normalised to range [0,1]) may therefore be written as:

$$L(\underline{c}) = \left((\underline{c} - \underline{\mu})^T \underline{\underline{U\Lambda U}}^T (\underline{c} - \underline{\mu})\right)^{\frac{1}{2}} \tag{3.16}$$

More precisely, taking into account $c_h = c_h \ \mathbf{mod}\ 1$ we use:

$$L'(\underline{c}) = \min(L(\underline{c}), L(\underline{c} + \begin{bmatrix} 0.5 \\ 0 \end{bmatrix})) \tag{3.17}$$

This produces a modified Mahalanobis field such as that of Figure 3-19, left (inset). Canny edges [12] are detected within this map to produce a set of binary edge pixels.

Faces vary in their shape; some tend toward elliptical forms whilst others are described as being more rectangular. In line with this observation we have chosen to model the face as a superquadric (equation 3.2) which encompasses all of these forms in a single parameterised framework. Superquadrics are fitted to the edge pixels using a Hough based search technique, which incorporates a novel implementation strategy to handle the large parameter space defining potential superquadrics. The reader is referred to Appendix A.1 for a full discussion of this fitting process. The fitting process results in a ranked list of 6-tuples, each corresponding to a potential location for a superquadric (Figure 3-19, left). Each 6-tuple contains a set of parameters: $[C_x,\ C_y,\ r,\ a,\ \theta,\ \alpha]$, which correspond to the 2D centroid location, scale, eccentricity, orientation and form factor respectively.

We take advantage of the natural symmetry of faces to assist in selecting the best fit superquadric from those returned. We quantify the symmetry (about the principal axis)

of the image region bounded by each candidate superquadric. For a given superquadric, the line of symmetry passes through point $(C_x, C_y)^T$ at angle θ degrees from the vertical. The reflected image $\underline{p}'$ of point $\underline{p}$ in homogeneous form may be obtained using the following affine transformation:

$$\begin{aligned} \underline{p}' &= \underline{\underline{T}}^{-1}\underline{\underline{R}}^{-1}\underline{\underline{F}}\underline{\underline{R}}\underline{\underline{T}}\underline{p} \qquad (3.18) \\ \underline{\underline{T}} &= \begin{bmatrix} 1 & 0 & -C_x \\ 0 & 1 & -C_y \\ 0 & 0 & 1 \end{bmatrix}, \ \underline{\underline{R}} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \underline{\underline{F}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

We perform the reflection on all pixels within the candidate superquadric, and compute the mean squared error (MSE) of the reflected and original images (Figure 3-19, middle). We score each candidate by multiplying the integer "evidence" for that superquadric (returned by the Hough search, see Appendix A.1), with unity minus its MSE. The highest scoring candidate is deemed to correspond to the facial region.

Locating features within the face

Facial features are located using a simple template based correlation scheme. Prior to processing, several eyes and mouths are manually segmented and registered upon one another to produce "average" feature templates. We used sixteen samples from the Olivetti face database [136] for this task; these faces are available only as greyscale images, and the template matching process thus correlates using only luminance data. The pixel data for each template is stored, along with the average triangle formed by the centroids of the two eyes and the mouth. This results in a rigid, planar template of the three features, which we attempt to register on to the facial region identified at run-time.

Registration is a two step process. First, the basis of the template is rotated to align with the ortho-normal basis of the superquadric bounding the facial region. Second, the template is subjected to translation $(T_x, T_y)^T$ and uniform scaling s in order to minimise the MSE between the template and the image data. Minimisation is performed using a Nelder-Mead search [114] to locate the optimal triple (T_x, T_y, s). Coarsely fitting shapes (circles for the eyes, rectangles for other features) are also drawn onto the template prior to processing. The vertices of these shapes form the initial control points for the active contours (snakes) which are relaxed on to salient features as per Section 3.4.1 (Figure 3-19, right), once the template has been fitted. The template also assigns preset equivalence class categories to the localised features, for example "eye", "nose" etc.

Note that we assume that this initial image contains a full frontal image of the face free from occlusion; though this image may be subjected to affine variations. This seems a reasonable constraint to impose given the photo booth application that motivates this case study.

3.5.2 Tracking the Isolated Salient Features

We can further automate the Cubist system by tracking identified salient features over consecutive frames of video. This implies that a temporal dependency must exist between source images, which has not been a constraint on the process until this point. However since the majority of source imagery is likely to be captured in the form of video, and this may be acceptable in the majority of cases.

The tracking process commences directly after the snake relaxation step for the initial frame (see previous subsection). We write $\underline{p}_i$ to represent the i^{th} of n inhomogeneous control points describing the spline fitted about a salient feature. We may write these points as a column vector to obtain a point $\underline{\phi}$:

$$\underline{\phi} = \left(\underline{p}_1^T, \underline{p}_2^T, ..., \underline{p}_n^T\right)^T \in \Re^{2n} \tag{3.19}$$

The problem of tracking a salient feature from one frame to the next is now reformulated to that of determining the mapping of the feature's bounding spline $\underline{\phi}$ to a new point $\underline{\phi}' = M(\underline{\phi})$ in the high dimensional space $\Re^{2n}$. If we assume all points on the feature boundary to be co-planar, this mapping $M(.)$ decomposes into a homography $\underline{\underline{H}}$ plus some additive term representing spatial deformation $\underline{s}$. Writing $\underline{\underline{P}}$ as a homogeneous representation of the points encoded in $\underline{\phi}$:

$$\underline{\underline{P}} = \begin{bmatrix} \underline{p}_1, \underline{p}_2, ..., \underline{p}_n \\ 1 \end{bmatrix} \tag{3.20}$$

we write the mapping as:

$$\underline{\underline{P}}' = \underline{\underline{HP}} + \underline{s} \tag{3.21}$$

If we assume that the object being tracked deforms very little from frame to frame, then all image-space deformation is due to viewpoint change. Under these conditions, homography $\underline{\underline{H}}$ well describes the mapping $M(.)$:

$$\underline{\underline{H}} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \tag{3.22}$$

where h_j is the j^{th} component of the homography $\underline{\underline{H}}$ (by convention, $\underline{\underline{H}}$ is normalised such that $h_9 = 1$; the transformation has eight degrees of freedom). We can therefore consider an eight dimensional subspace in $\Re^{2n}$, within which points corresponding to valid splines (and so tracked features) will approximately lie[3]. Deviation from this space corresponds to feature shape deformation during tracking. We assume the manifold of valid splines to be locally linear and so well approximated by a hyper-plane (this assumption is justified momentarily). The bases of this plane are obtained by computing the Jacobian of $M(.)$. Applying a Taylor expansion of 1^{st} order to $M(.)$ we obtain:

$$M(\underline{\phi} + \mathrm{d}\underline{\phi}) = M(\underline{\phi}) + \underline{\nabla}^T_{M(\underline{\phi})}\mathrm{d}\underline{\phi} \tag{3.23}$$

where $\underline{\nabla}^T_{M(\underline{\phi})}$ is the gradient of $M(\underline{\phi})$ at $\underline{\phi}$. Under our assumption that $M(.)$ varies only by homography, then $\underline{\nabla}_{M(\underline{\phi})}$ may be written as:

$$\underline{\nabla}_{M(\underline{\phi})} = \begin{bmatrix} \frac{\partial \underline{\phi}_1}{\partial h_1} & \frac{\partial \underline{\phi}_1}{\partial h_2} & \cdots & \frac{\partial \underline{\phi}_1}{\partial h_8} \\ \frac{\partial \underline{\phi}_2}{\partial h_1} & \frac{\partial \underline{\phi}_2}{\partial h_2} & \cdots & \frac{\partial \underline{\phi}_2}{\partial h_8} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \underline{\phi}_{2n}}{\partial h_1} & \frac{\partial \underline{\phi}_{2n}}{\partial h_2} & \cdots & \frac{\partial \underline{\phi}_{2n}}{\partial h_8} \end{bmatrix} \tag{3.24}$$

where $\frac{\partial \underline{\phi}_i}{\partial h_j}$ denotes the shift of the i^{th} control point under a small change of the j^{th} component of the homography.

The basis of the valid subspace corresponds to the eight columns of $\underline{\nabla}_{M(\underline{\phi})}$, scaled by the reciprocal of the square of their respective L_2 norms. This process compensates for the greater influence over motion that some homography components (e.g. projections) hold over others (e.g. translations) given similar numerical variation. The remaining null space $\Re^{2n-8}$ accounts for arbitrary shape deformations of the tracked spline. Generating small normal variate offsets from $\underline{\phi}$ within the homography space basis set generates "similar" splines, related by homography to the spline specified by $\underline{\phi}$ (see Figure 3-20). Notice that as projections are cast further from the original contour they tend away from homographies and begin to exhibit shape deformations. This is because our linear approximation is only local to $\underline{\phi}$, and deteriorates as we move further from $\underline{\phi}$ so digressing from the space of valid contours into the shape deformation (null) space.

A two stage process is employed to track features. First we determine the homography

[3] In fact, our $\Re^{2n}$ parameter space was chosen specifically because of its support for shape change due to homography. Such support is not generally present for any parameter space; consider, for example, the 6D parameter space of the superquadric (Appendix A.1).

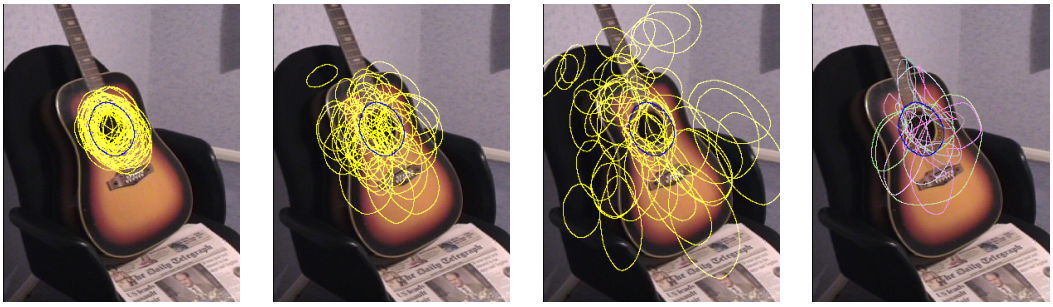

Figure 3-20 The bounding spline of a salient feature (the guitar hole) after snake relaxation (dark blue). The three leftmost images show variation in homography space using normal variates of standard deviation 20, 50, and 100 (from left to right). Shape deformation becomes apparent with greater standard deviation as the local linear approximation deteriorates. By contrast the rightmost image shows variation in the null (arbitrary shape deformation) space.

that best maps the feature from one frame to the next. In this context the "best" mapping corresponds to a minimisation of the mean squared error (MSE) $E(.)$ between the RGB colour values of pixels bounded by the spline in the current frame $I_t(.)$ and those bounded by the putative spline location in the next frame $I_{t+1}(.)$:

$$E(M_i(.); \underline{\phi}, I) = \frac{1}{N} \left| I_{t+1}(M_i(\underline{\phi})) - I_t(\underline{\phi}) \right|^2 \tag{3.25}$$

where N is the number of pixels bounded by spline $M_i(\underline{\phi})$. $M(.)_i$ is a putative mapping (homography) obtained by local stochastic sampling in the valid subspace as previously described — we use a random variate centred at $\underline{\phi}$, with a preset standard deviation σ. Choice of σ effectively determines the maximum inter-frame distance an object can move, such that we may continue to track it; note that σ can not be too large, or the locally linear approximation to the homography breaks down. Values up to around 50 are reasonable for video frame size images; lower values are possible for tracking slower movements. This process yields an approximate grouping contour which is used as the initial position for a snake which is iteratively relaxed to account for any shape deformation in the feature (as per the method of Section 3.4.1). We perform this relaxation to take into account shape deformations, which we do not wish to track explicitly. This is due to their high dimensionality and unconstrained nature, which has been shown to cause unstable tracking in the absence of a predetermined, specific motion model [84] (which we did not wish to introduce for reasons of generality).

This approach does not yet take into account the possibility of occlusion — however features can become occluded when subjected to large camera viewpoint changes. We handle this problem in two ways. First, we may choose to track individual features (as in Figure 3-20), or the entire set of features simultaneously under the assumption that all features are co-planar in the world. This latter approach holds the advantage

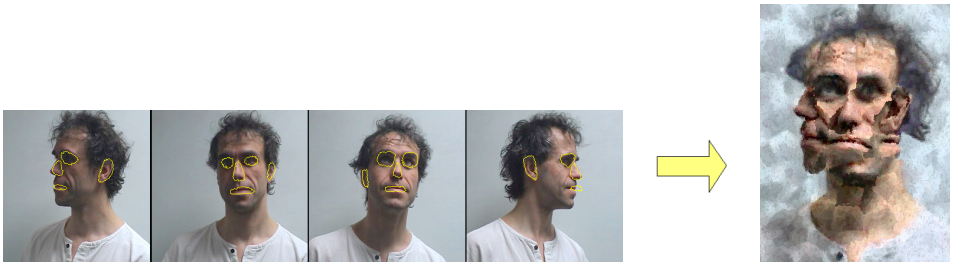

Figure 3-21 Four frames of a source video after image registration. Features have been tracked as a planar group, and each active contour subsequently relaxed to yield a salient feature boundary (shown in yellow). The corresponding Cubist painting is on the right.

that if one salient feature, say an eye, rotates away from the view of the camera, the other features may still be tracked. This yields a robust common homography under which all features are assumed to have moved. We have found this approach to work acceptably well for faces (Figure 3-21). Second, if the MSE for a feature (equation 3.25) rises above a preset upper threshold, then that feature is deemed to be occluded. In such cases, the feature is not to be sampled from the current frame, to prevent garbled image fragments being forwarded to the composition engine in place of the occluded feature. The equivalence classes attached to a feature persist over time during tracking.

We stated earlier that a locally linear approximation to the homography space was satisfactory for our application. We have shown this to be true by examining the 2^{nd} order Taylor expansion of $M(.)$. This results in eight second order Jacobians, which we observe have determinants close to zero in all but the cases in which h_7 and h_8 vary. Thus the valid space is approximately linear except for the projective components of the homography. If we make the intuitively valid assumption that that change of viewpoint is small (relative to other homography components) from one frame to another, then a linear approximation to the space is justified.

The decomposition approach allows us to track robustly since we restrict the possible deformations of the contour to those caused by a change of viewpoint, which we assume accounts for the majority of deformation in the 2D scene projection. However, there are further advantages gained through decomposition of motion into a homography plus some general shape deformation. As each transform $\underline{\phi}' = M(\underline{\phi})$ is computed, we calculate a motion vector $\underline{\phi}' - \underline{\phi}$ in the valid subspace $\Re^8$. Based on the line integral of the path traced in this space, we are able to quantify the magnitude of change of viewpoint. Since we approximate the space to be locally linear we are justified in using a Euclidean distance to compute this inter-frame distance; however the path integral approach is necessary over larger time periods, since we have shown the space to be globally non-linear. When this accumulated distance rises above a certain threshold,

we sample the current frame from the video for use in the composition. Thus the selection of frames for composition is also automated (subject to the user setting a suitable threshold for the integral function).

Future work might attempt to further decompose motion into the various components of the homography to estimate how the feature has translated, rotated etc. Currently the Cubist algorithm performs a translational image registration using distinct colour signatures within the image. Using the homography to determine an inverse translation may allow us to register images regardless of their colour characteristics, and we predict that this may improve the generality of the algorithm.

3.6 Summary and Discussion

We have argued that the paradigm of spatial low-level processing limits AR in two ways. First, that quality of rendering suffers since magnitude of high frequency content, rather than the perceptual importance, of artifacts governs emphasis during rendering. We have shown that addressing this limitation demands global image analysis, rather than the spatially local approach so far adopted by AR. Second, that the spatially local nature of processing limits AR to low level, stroke based styles. We argued that the synthesis of compositional forms of art, such as Cubism, can not be achieved without processing images at a spatially higher level than that of local pixel neighbourhood operations.

In this chapter we introduced a global salience measure to AR, to determine the relative importance of image regions. We applied this measure to propose two novel AR algorithms, which respectively addressed each of the AR limitations identified in the previous paragraph. The first was a single-pass AR algorithm capable of rendering photographs in a painterly style reminiscent of pointillism. This algorithm adaptively varies the emphasis in a painting to abstract away non-salient detail, and emphasise salient detail. The second was an algorithm capable of producing compositions in a style reminiscent of Cubism. Uniquely, this algorithm made use of salient features (eyes, ears, etc.) as the atomic element in the painting, rather than the low-level stroke. The two algorithms respectively demonstrate how a spatially higher level of image analysis can improve the aesthetic quality of renderings (more closely mimicking the practice of human artists), and extend the gamut of AR beyond stroke based rendering to encompass compositional artistic styles such Cubism.

There are a number of directions in which this work might be developed further. We have shown the introduction of a global salience measure can remove limitations im-

posed by spatially local nature of current AR. Although this rarity based salience measure is new to computer graphics, it is quite simplistic. The definition of image salience is highly subjective and context sensitive. Consider a snap-shot of a crowd: in one scenario a particular face might be salient (for example, searching for a friend); in another scenario (for example, crowd control), each face might hold equivalent salience. The development of image salience measures is an area of considerable interest in Computer Vision, and no doubt other global salience measures might be substituted for our own — the loose coupling between the salience measure and rendering steps facilitates this. Indeed, in Chapter 4 we make use of a more subjective, user trained measure of salience to drive a more sophisticated salience adaptive painterly rendering process, which follows on from this work.

The Cubist rendering algorithm is an illustration of the potential expansion of AR's gamut of artistic styles that can be achieved by considering spatially higher level features within a scene. However to achieve this higher level of segmentation we must impose a restrictive model upon the scene, removing the need for interaction, at the cost of generality. It is unfortunate that contemporary Computer Vision techniques limit full automation to only a few well studied cases. However, interactive grouping typically takes less than one minute of user time, and so we are content with our method as a compromise between a general system and automation. As regards the "Personal Picasso" proof of concept system, although the face localisation algorithm is reasonably robust, the location of facial features themselves leaves something to be desired. Likewise, the tracker is adequate but could be improved to be more robust to occlusion. The implementation of a more sophisticated facial feature location and tracking system for the Cubist renderer is a live BSc. project at Bath.

Although Chapter 4 serves as a continuation of the pilot painterly rendering algorithm presented in this Chapter, there are a number of interesting directions the Cubist rendering work might take. The depiction of movement within a static image is a unique contribution to AR, but may hold further applications. For example, the production of static "thumb-nail" images to help summarise and index video content. We might consider undertaking a compositional analysis in order to more aesthetically place our high level features, and progress yet further toward emulating Cubism; however we believe such an analysis is a considerable challenge that is not necessary to demonstrate the synthesis of Cubist-*like* renderings are made possible through higher level spatial analysis. We might revisit the way in which we apply paint, so that it appears more in the tradition of a particular artist, but there is no compelling reason to focus our work in such a way at this stage: the manipulation of high level features is the only necessary step to producing images that can be classified as "Cubist" or, at least, "Cubist

influenced".

A higher level of analysis still, might be applied to extract salient features from image; perhaps a full 3D scene reconstruction and re-projection from novel perspectives to generate alternative Cubist-like styles. Perhaps alternative global analyses of the image might generate aesthetically pleasing abstract artwork. For example, one might investigate use of the Hough transform to identify target shapes within an image, taken from a user defined "shape library". The subsequent rendering of those shapes may provide a basis for synthesising alternative abstract artistic styles. Such possibilities lend further credence to our argument that higher level spatial analysis opens the doorway to a wide range of otherwise unobtainable artistic rendering styles.