IMAGE RECOGNITION USING MAXIMAL CLIQUES OF INTEREST POINTS

Fred Stentiford[1] *& Ade Bamidele*[2]

[1]University College London, Electronic & Electrical Engineering Dept, Gower St, London, UK
[2]Nokia UK Limited, Nokia House, 10 Great Pulteney St, Westminster, London, UK
f.stentiford@ee.ucl.ac.uk & ade.bamidele@nokia.com

ABSTRACT

This paper describes a method for extracting interest points in images and using interrelated groups or cliques to recognize structure common to pairs of images. Feature measurements are commonly selected intuitively and work well on data that are thoroughly understood. This approach avoids the use of global features and relies upon candidate interest points and their relationships with each other. The method is applied to photos of movies posters and the results are compared with those achieved with SIFT key points.

Index Terms— Pattern recognition, Image recognition.

1. INTRODUCTION

Extracting features from images and videos is central to problems of identification, categorization, tracking and retrieval where semantically relevant features are necessary to obtain recognition. However, performance on recognition problems involving many diverse and unpredictable patterns is rarely satisfactory using preselected features. A feasible approach to these problems has been to automatically locate and characterise interest points in images that are perceptually important and therefore likely to play a significant role in a recognition task. Furthermore the relationships between interest points may also be used to measure similarity.

2. BACKGROUND

Harris points are determined by locating points that have large gradients in all directions at a predetermined scale [1]. Later developments require the principal curvatures at interest points to be greater than some minimum [2].

The Scale Invariant Feature Transform (SIFT) generates large numbers of interest points which possess values of scale and orientation and are used for object recognition [3]. The method applies a series of difference-of-Gaussian functions reflecting different scales to the image and locates interest points at extrema in these transformed images. A mechanism is described for matching key points between images that is used for object recognition. Lindeberg selects scale by looking for maxima in the Laplacian of Gaussian applied to (x, y, scale) [4]. The SUSAN detector uses a circular mask of fixed radius to extract local structure. The area of pixels within the mask that match the centre pixel reaches minima over corners and edges [5]. Kadir et al find interest points at positions of maximum entropy over scale [6]. This work assumes that saliency is directly related to complexity and noise. Mikolajczyk et al extend Lindeberg to generate key points whose descriptors are affine independent and can be matched under affine transformations [7]. Belongie et al [8] define the context of an image sample point as its relationship to all other sample points in the image and uses this information to obtain recognition. Strategies are invoked to reduce the variability between instances in the same category. Leordeanu et al make use of pairwise relationships between points and build a composite graphical model from training images in the same class [9].

Ogawa [10] uses a Delaunay triangulation of interest points in pairs of images before extracting maximal cliques of points from matching in their properties as well as their angular relationships. The triangulation partitions candidate matching points and reduces computation but restricts the possible clique structures that may be extracted. Bolles et al [11] employ a similar mechanism to locate industrial parts using manually corners and holes as interest points. Horaud et al [12] extract cliques from graphs representing straight edges and their relationships to obtain a matching for stereo correspondence.

The approach taken in this paper makes use of interest points and their relationships with each other. The similarity of pairs of images is measured by the extent that cliques of matching interest points possessing the same angular relationship with each other are present in both images.

3. INTEREST POINT GENERATION

It is important when extracting interest points for the purposes of recognition that any assumptions made about the data do not exclude information that will be potentially useful for discrimination. Colour as perceived in the human visual system may vary independently in three dimensions as typically reflected in the Lab colour space [13]. Each colour channel may therefore contain discriminating information not present in the others and it follows that interest points can potentially exist in different positions depending on the colour channel.

Interest point generation begins with the random selection of 1000 ordered pair of pixels $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ that differ in a colour channel *a* by a threshold δ_a in the image *I*. Such pairs of pixels would be very common in a noisy image and would not reflect local structure very well. The significance would be greatly enhanced if other pixel pairs in the vicinity behaved in the same manner. The local agreement on colour difference is therefore measured in order to assess whether the pixel pair should be retained.

Pixels $\boldsymbol{x}_k$ within a circle C centre $\boldsymbol{x}_1$, radius $r_1 = |\boldsymbol{x}_1 - \boldsymbol{x}_2|$, are compared with the pixel $\boldsymbol{x}_2$ using the colour channel *a*. The *significance* S_a of $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is given by

$$S_a = \left| \sum_{\boldsymbol{x}_k \in C} \boldsymbol{x}_1 - \boldsymbol{x}_k \right| \Big/ |\boldsymbol{x}_1 - \boldsymbol{x}_2|^2 \qquad (1)$$

where $|u_a^{x_k} - u_a^{x_2}| < \delta_a$ and $u_a^{x_k}$ is the value in the colour channel *a* at pixel $\boldsymbol{x}_k$.

The magnitude of S_a will be a maximum if $\boldsymbol{x}_1$ lies on a straight edge bounding the colour of $\boldsymbol{x}_2$ from others. It will be zero if the circle C is all one colour. The orientation of the local gradient is given by

$$\theta_a = \tan^{-1}\left\{ \left(\sum_{\boldsymbol{x}_k \in C} \boldsymbol{x}_1 - \boldsymbol{x}_k \right)_y \Big/ \left(\sum_{\boldsymbol{x}_k \in C} \boldsymbol{x}_1 - \boldsymbol{x}_k \right)_x \right\} \qquad (2)$$

Equation (1) measures the extent of mismatch between pixels of similar colour to $\boldsymbol{x}_2$ and others on the opposite side of $\boldsymbol{x}_1$.

Points with the highest value of significance S_a within a radius r_2 are retained thereby providing interest points covering the image regardless of local contrast.

4. MATCHING AND RECOGNITION

Simply comparing the highest ranking interest points between two patterns for orientation and the colours of the respective pixel pairs, takes no account of the positional relationships between the interest points. However, this may be overcome by seeking matches between pairs of interest points where the orientations of the relative positions are similar in each pattern.

Let $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ be interest points in colour channel *a* with orientations θ_i and θ_j. Let $d_{ij} = |u_a^{x_i} - u_a^{x_j}|$.

$\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ match if $d_{ij} \le \delta_a'$ and $|\theta_i - \theta_j| \le \varepsilon_1$ (3)

where ε_1 is a threshold on angle difference.

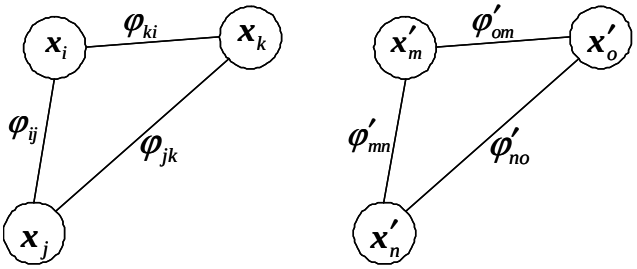

Figure 1. Matching cliques of size 3

Let $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ be interest points from image 1 and $\boldsymbol{x}'_m$ and $\boldsymbol{x}'_n$ be interest points from image 2. The interest point pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ matches the pair $(\boldsymbol{x}'_m, \boldsymbol{x}'_n)$ if $\boldsymbol{x}_i$ matches $\boldsymbol{x}'_m$ and $\boldsymbol{x}_j$ matches $\boldsymbol{x}'_n$ and

$$\frac{(\boldsymbol{x}_j - \boldsymbol{x}_i) \bullet (\boldsymbol{x}'_n - \boldsymbol{x}'_m)}{|\boldsymbol{x}_j - \boldsymbol{x}_i| * |\boldsymbol{x}'_n - \boldsymbol{x}'_m|} \ge \lambda \qquad (4)$$

The inner product in equation (4) constrains the difference in slopes between the pairs of points in each image to be less than a certain angle ε_2

$$|\varphi_{ij} - \varphi'_{mn}| < \varepsilon_2 \text{ where } \lambda = \cos\varepsilon_2 \qquad (5)$$

The matching of the pairs of interest points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ and $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ has greater reliability if the pair $\boldsymbol{x}_k$ and $\boldsymbol{x}_i$ also match as this shows that all three matching points are in the same relative angular position in each image. Graphs with vertices representing matching points $\{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k\}$ and $\{\boldsymbol{x}'_n, \boldsymbol{x}'_m, \boldsymbol{x}'_o\}$ and connected by angular relationships φ_{ij} form cliques of size 3 present in both images (Fig. 1). More generally the detection of *p* such points in each image forming cliques will provide greater recognition reliability for larger values of *p*. The measure of similarity between two images used in this paper is defined as the number of distinct interest points that are members of cliques of size > N summed over the three colour channels.

The relationship between points is not dependent upon their separation or absolute position and therefore the similarity measure is translation and scale invariant. However, the relationship between points is only partially orientation invariant, but may be made so by matching relative angles within each clique.

5. RESULTS

The similarity measure was evaluated by comparing two sets of 11 photos (176x130) of movie posters; one higher quality set (R1-R11) and the other a lower quality set of photos of the same posters (C1-C11).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
R1	53	0	0	0	0	6	5	0	0	0	6
R2	0	31*	6	0	0	0	0	0	0	0	0
R3	6	0	17	0	0	0	0	9	0	0	10
R4	0	18	0	42	0	5	0	8	17	0	0
R5	0	0	0	0	30	0	0	0	0	0	5
R6	0	17	0	0	0	21	0	0	0	0	10
R7	0	0	0	0	0	0	18	0	0	0	0
R8	8	12	0	11	0	8	0	33	5	0	20
R9	0	0	0	0	0	0	0	0	49*	0	0
R10	0	0	0	0	0	0	0	14	5	40*	0
R11	10	13	7	0	0	5	0	0	5	0	50*

Table 1. Numbers of matching interest points

100 interest points were generated in each of the three 8 bit Lab colour channels for each image with $(r_1, r_2) = (5, 4)$ pixels and $(\delta_L, \delta_a, \delta_b) = (6,6,6)$. Matching cliques with $N>5$ were extracted for each pair of images and the similarity scores shown in Table 1 with $(\delta'_L, \delta'_a, \delta'_b) = (21,21,21)$ and $(\varepsilon_1, \varepsilon_2) = (27^o, 15^o)$. In each case the lower quality image (Cn) produced the highest score when matched against the higher quality image (Rn) of the same poster. Some examples of cliques are shown in Figures 2-5. Green circles show the location of interest points and red radial lines the local orientation.

The extraction of maximal cliques is an NP-complete problem and requires exponential time to solve. However, the size of matching graphs only grows when images possess strong similarities and in these cases the computation can be limited without losing much recognition capability. Entries in Table 1 marked with an asterisk have triggered the application of a computation limiting algorithm and may be underestimates of the actual scores.

The SIFT method of key point extraction and matching was applied to the same problem and the numbers of matching key points between pairs of images summed across the three Lab colour channels is shown in Table 2. No matching key points were found in C2 and only a few in C7 possibly because these images possess low contrast.

6. DISCUSSION

The low values of δ_i allows interest points to be detected in areas of relatively low contrast. At the same time the high values of δ'_i allows a large tolerance to variations in lighting and reflectance during recognition (Fig. 4 & 5). The value of ε_1 allows interest points to be matched with only an approximate correspondence in orientation. In a similar fashion ε_2 provides considerable leeway when matching the direction of pairs of points between images and provides some immunity to perspective distortions (Fig. 2 & 4). Results show that providing the matching cliques are sufficiently large, significant variation in individual pattern parameters can be tolerated. This means that confident recognition is more likely despite highly variable and distorted images.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
R1	17	0	0	0	0	0	1	0	0	0	0
R2	0	0	0	0	0	0	1	0	0	0	1
R3	0	0	39	0	0	0	0	0	0	0	0
R4	2	0	1	11	0	0	1	1	0	2	0
R5	0	0	0	1	59	0	1	0	0	0	0
R6	0	0	0	1	2	19	0	0	0	0	0
R7	0	0	0	0	0	0	3	1	1	0	0
R8	0	0	1	0	2	0	0	8	0	0	0
R9	0	0	0	0	0	0	0	0	17	0	0
R10	0	0	0	0	0	0	0	0	1	22	0
R11	1	0	0	0	0	0	0	0	0	0	16

Table 2. Numbers of matching SIFT key points

7. CONCLUSIONS

This paper has described an approach to the generation and selection of interest points and their formation into cliques for use in image recognition. The computational demands for clique extraction rise when strong similarity is present and may be limited on these occasions without damaging performance. Future work will focus on orientation independent clique extraction with larger datasets and more interest points.

8. REFERENCES

[1] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," Proc 4th Alvey Vision Conference, pp. 147-151, 1988.

[2] J. Shi and C. Tomasi, "Good features to track," IEEE Conf on CVPR, 1994.

[3] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. of Computer Vision, vol. 2, pp. 91-110, 2004.

[4] T. Lindeberg, "Feature detection with automatic scale selection," J. Computer Vision, vol. 30, no. 2, pp. 79-116, 1998.

[5] S.M. Smith and J.M. Brady, "SUSAN – a New Approach to Low Level Image Processing," Int J. of Computer Vision, vol. 23, no. 1, pp. 45-78, 1997. (Patent GB2272285).

[6] T. Kadir and M. Brady, "Saliency, Scale and Image Description," Int. J. of Comp Vis, vol. 45, no. 2, pp. 83-105, 2001.

[7] K. Mikolajczyk et al, "Scale and Affine Invariant Interest Point Detectors," Int. J. of Comp Vis, vol. 60, no. 1, pp. 63-86,2004.

[8] S. Belongie, et al, "Shape matching and object recognition using shape contexts," IEEE Trans PAMI, vol 24, no. 24, pp 509-522, 2002.

[9] M. Leordeanu, M. Hebert and R. Sukthankar, "Beyond local appearance: category recognition from pairwise interactions of simple features," CVPR 2007.

[10] H. Ogawa, "Labeled point pattern matching by Delaunay triangulation and maximal cliques," Pattern Recognition, vol 19, no 1, pp 35-40, 1986.

[11] R. C. Bolles and R. A. Cain, "Recognising and locating partially visible objects: the local-feature-focus method," Int J Robotics Research, vol 1, no 3, pp 57-81, 1982.

[12] R. Horaud and T. Skordas, "Stereo correspondence through feature grouping and maximal cliques," IEEE Trans PAMI, vol 11, no 11, pp 1168-1180, 1989.

[13] Supp. No. 2 to CIE Publication No. 15, Colorimetry, 1976.

Figure 2. Matching 7-cliques in the *b* channel in images R4 and C4†

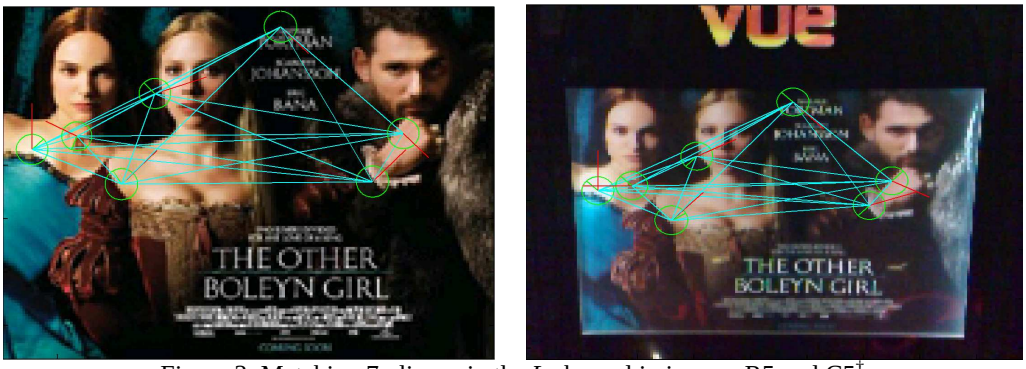

Figure 3. Matching 7-cliques in the *L* channel in images R5 and C5†

Figure 4. Matching 6-cliques in the *L* channel in images R2 and C2†

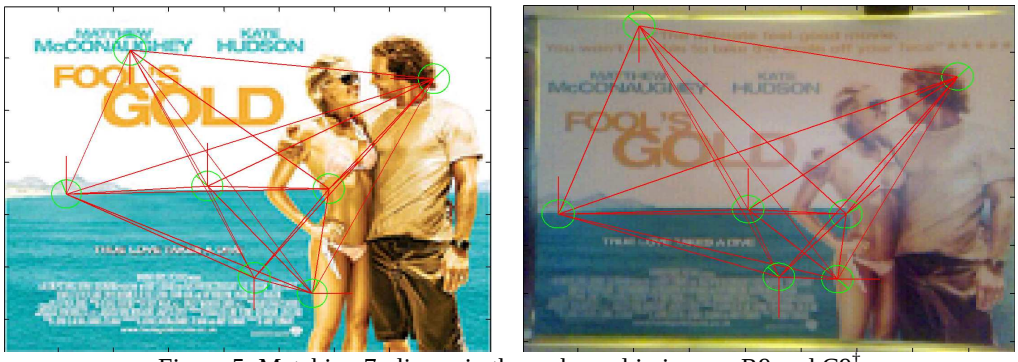

Figure 5. Matching 7-cliques in the *a* channel in images R9 and C9†

† Images are courtesy of Nokia Point&Find for illustration purposes only.