Synthetic Image of Multiresolution Sketch Leads to New Features

Georgii Khachaturov and Rafael Moncayo-Muños
Unuversidad Autónoma Metropolitana
xgeorge@tarunz.org

Abstract

A new approach to construction of robust features is proposed and applied to an instance of the correspondence problem. The main idea is to construct a synthetic image by a multiresolution sketch (MS) of an image and involve it into extraction of the invariants. The MS is constructed by processing the image with a scalable detector of the semi-local 1D-elements. Then, a synthetic image is constructed with all elements of the MS. Local maxima of the first and second derivatives of the synthetic image along discrete curves of the MS lead to some singular elements represented by the points of a 4D manifold. It turns out that a representative subset of the singular elements is stable. To prove that, the pair-wise correspondence between subsets of singular elements of two shots of a film was established experimentally by a consistency technique, which, unlike past approaches, does not involve epipolar constraints.

1. Introduction

1.1. Objective. The robust features able to form a dense network in the image frame are important for different application of Computer Vision (CV).

In principle, the 1D features would have some important advantages in use over the 0D features. However, while using a single-scale 1D detector, the construction of a 'good' discrete set of the 1D features is impossible because, given an extended 1D object, it is impossible to distinguish its separate 1D elements. What could destroy this obstacle is to take into consideration the dynamical changes of the 1D objects found with a variable-scale detector.

The presented work is aimed at the construction of 1D features with the 'good' properties, using scaling.

1.2. A new tool for the correspondence problem. Anticipating presentation of the proposed technique of early and intermediate image processing, we illustrate its impact for the correspondence problem, which is a well-known problem of CV [19].

In a primitive form, the correspondence problem is as follows: Given U_1 and U_2 as two discrete sets formed by the features of a certain kind extracted from two different images, determine two subsets $R \subseteq U_1$ and $S \subseteq U_2$ and a one-to-one correspondence between them, $R \leftrightarrow S$, to be optimal in a certain meaning.

In most instances, U_1 and U_2 are formed by many low quality elements. So the problem is to determine some 'inliers' to fit a certain model and separate them from the 'outliers'. Usually inliers are formed by a group of features consistent with a mapping that relates the images, so the goal is to determine the group and the mapping simultaneously.

Instances of the correspondence problem vary mainly in the following issues: (*i*) the number of elements in the sets U_1 and U_2, which presents a combinatorial constraint on the computational complexity of the algorithm; (*ii*) configuration space that contains the features; (*iii*) stability of the features in response to variation of the image content; (*iv*) precision of the features; (*v*) class of real world objects represented in the images and marked by the features.

Novelty of the early and intermediate level processing presented in this paper can be illustrated by its effect on the above issues:

(*i*) dimension of the configuration space of the new features is equal to four with some three parameters measured with a high accuracy and the remaining one with a low accuracy, versus two-dimensional features in most conventional techniques;

(*ii*) a single correspondence $\{r_i, s_j\}$, $r_i \in U_1$, $s_j \in U_2$, yields an approximation to matching of small image fragments related to r_i and s_j, respectively, versus at least several correspondences required for that in past approaches;

(*iii*) the class of real objects to match in different shots now includes the flexible objects of a complex configuration subjected to dynamical changes, versus typical for past approaches the rigid objects formed as a piecewise linear structure.

1.3. Description of results. A scalable detector of semi-local 1D elements is applied with different scales to construct a multiresolution sketch (MS) of the input

image. The MS is a discrete structure formed by the detected elements for different scales and each element can be represented as a point in a 4D configuration space. Then, a synthetic image is constructed with the MS. A joint processing of the MS and the synthetic image leads to extraction of some singular elements related to simple geometric invariants.

Any singular element g is used as the nucleus to construct a sub-structure G in the MS. A correspondence between g and g_1 from two different MSs induces a matching of the respective G and G_1. This matching leads to a good similarity measure for potential correspondences.

Using this similarity measure and a kind of Hough transform, an algorithm for solving the correspondence problem was developed.

Experiments show that this algorithm successfully finds the pairs of corresponding elements for two shots of a film with a complex content and dynamics, which proves relevance of the new features.

1.4. Related works. The proposed approach is a combination of different techniques. The choice of a particular component in the scope of each technique is not very stiff. The following brief review compares each component of the combination with the alternative ideas of a correspondent technique.

Detection of 1D elements. For actual goals, an extended version of a scalable detector called "three frequencies method" (3FM) [9] is applied. It is a non-linear detector that deals with gray scale images.

The features detected by the 3FM belong to a wider class than the class represented by traditional edges. For instance, either a traditional edge or a fragment formed by several parallel strips of different widths is regarded by the 3FM as a single 1D element of the extended class. The term `gedgel` (generalized edge element) stands for the objects of this extended class.

The following properties of 3FM are substantial for us: (*i*) the extended class of detected objects, (*ii*) a higher accuracy (compared to conventional edge detectors) of the detected elements orientation, (*iii*) possibility of scaling.

Typically, an edge detector [1, 6, 10, 11, 14, 16, 18] seeks (in a small image fragment) the revelation of a specific property of a particular functional prototype ('step', 'roof', 'impulse', 'ridge', and 'ravine'). Unlikely, the 3FM deals with a mathematical property common for all imaginable prototypes.

For color images, edge detection is a more complex problem [17], mainly because the meaning itself of a color edge is still in the processes of understanding.

*Extraction of stable objects***.** For some applications, stability of objects extracted by pre-processing is even more important than their interpretability.

Interest points [3] represent a kind of the 0-D stable objects. Depending on context, the meaning of interest point may vary combining the following options: distinctness, invariance, stability, uniqueness and interpretability. The configuration space that contains the features represented by interest points coincides with the image frame, so it is a 2D space.

Snakes [8] are stable (global) objects detected by minimization of an energy functional that strengthens distinctness of individual fragments of the objects by *a priori* knowledge about their 'collective behavior'.

Snakes or blobs may be, or may be not related to a specific interpretation. However the stability of these objects in response to variation of the image forming factors is a really important property.

A recent example of the stable objects is presented by 'maximally stable extremal regions (MSER)' [13]. MSERs are used for solving a particular kind of correspondence problem called 'wide base-line stereo'.

The present approach introduces the MS as a kind of multi-scale space, widely used in CV at least from the works [11, 12]. The MS is formed as an ordered set of slices corresponding to different scales. For each slice, some curves are tracked (applying the 3FM), then merged into larger structures that belong to a smallest class of potentially distinguished objects in the slice.

The structures are unstable (as it should be for any kind of the non-model-based edge linking) and have no specific interpretation. However, a joint processing of such structural objects from different slices leads to a novel class of stable 4D features. The idea is to scan slices of the MS and measure torsion of the main direction of the structures as function of the scale; the features are detected in response to a high torsion.

Correspondence problem. This topic is included into the paper to show potential of the new features. Our solution is far from to be a final version because it explores only a small part of the information conveyed by the features, and is not studied deeply yet.

Two widely known ideas are combined in the proposed algorithm for solving the correspondence problem: search of local maxima for a kind of Hough transform, and a modification of RANSAC.

In graph theory, the analogue of the correspondence problem is called the matching problem [15]. *Matching* in a graph is a set of edges, no two of which share a node. Given a graph, the matching problem is to find a maximum matching. This problem has a natural modification called the weighted matching, if different weights are attributed to edges of graph.

The correspondence problem in CV involves some geometric constraints in addition to a graph structure and weights as for the matching problem. It has links to different areas, such as consistent labeling, Markov fields, and others [4].

In the last years, a considerable advance [5] in the correspondence problem was done with RANSAC [2], which is a probabilistic approach that finds a mapping between images as the best one from a constrained class of mappings. Mainly, the epipolar constraints were regarded so far. While applying RANSAC for other kind of features, the class of real objects for all experiments published so far [e.g.5, 13] was reduced to the rigid objects formed by plane and linear structures.

In contrast, it turns out that the features presented in this paper lead to a solution of the correspondence problem for a much wider class of real objects subjected to complex dynamical changes.

It is an old idea to involve a similarity measure for the potential pairs of corresponding elements into solving the correspondence problem. For example, it was applied in [3] to determine correspondences between interest points. There, it was based on a functional that deals with two small image fragments around interest points. Enlargement of these fragments can destroy adequateness of the similarity measure.

In contrast, for the actual technique, the fragments involved implicitly into construction of a novel similarity measure are considerably larger than for interest points, which leads to a higher robustness.

2. The 3-frequencies method (3FM)

Given image I represented as intensity function, I_V denotes below I reduced to a rectangle V in the image frame; $I_V(u,w)$ denotes value of I_V represented in own coordinate system of V for the axe $\boldsymbol{u}$ and $\boldsymbol{w}$ parallel, respectively, to sides $\boldsymbol{a}$ and $\boldsymbol{b}$ of rectangle V (Fig. 1A)**.**

Definition 1. In this notation, I_V is said to be *gedgel* parallel to the side $\boldsymbol{b}$ if and only if it can be represented as $I_V(u,w) = \lambda + \mu\varphi(u)$ for some real constants λ, μ and a real function $\varphi(u)$.

Remark 1. Notice that one comes to the traditional definition of edge by substituting the indefinite $\varphi(u)$ of Definition 1 with 'a certain prototype function $\varphi(u)$' and reducing the options for prototypes to the typical set of step-function, impulse-function, 'roof', etc..

For a known $\varphi(u)$, one may involve its concrete specific into detection, but it is not so for Definition 1.

Terminology note. There are three options for interpretation of any kind of objects detected in images (for example, edges): (*i*) Image fragment that satisfies a specific property; (*ii*) Feature assigned to a detected object; (*iii*) Arbitrary point in configuration space of the objects. In the rest of paper, we will use the term '*gedgel*' not only in the meaning of Definition 1, which corresponds to the first option, but in all three meanings, making special comment in cases when this strategy can lead to misunderstanding.

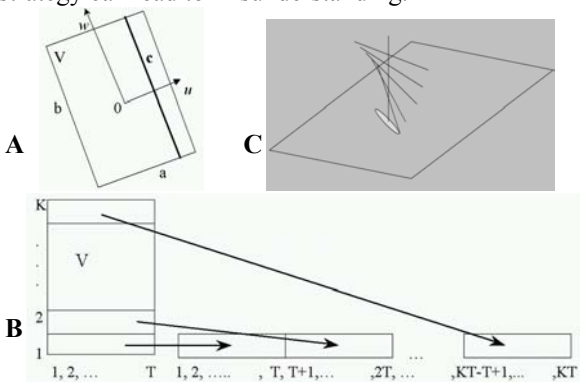

Figure 1. A: An illustration to Definition1. B: Unfolding the rectangular lattice in V into a string; the lattice in V is depicted by rows of length T, parallel to the side $\boldsymbol{a}$; columns have length K and are parallel to the side $\boldsymbol{b}$. The length of discrete line that unfolds V is KT. **C: An illustration to Section 4; a geometric invariant to be detected as singular gedgel**: For the sun close to zenith and a semi-transparent surface of the shape depicted in figure, a spot of darkest shade will stably persist close to the depicted ellipse.

Detection algorithm. The definition of gedgel implies periodicity of a function of one variable that unfolds all rows of the lattice inside V into one string (Fig. 1B). Period of this function is equal to the size of the rows, known in advance. Hence, the test on presence of a gedgel inside a rectangle window is equivalent to the test on periodicity of the corresponding unfolded function of one variable. The 3FM is such a test. It deals with three functionals denoted $S^0{}_{V\alpha}$, $S^1{}_{V\alpha}$, y $S^{-1}{}_{V\alpha}$ and defined for $n = -1,0,1$ as

$$S^n_{V_\alpha} = \frac{1}{KT} \left| \sum_{t=1}^{t=KT} f_{V\alpha}(t)\, e^{-i(1+n/K)2\pi t/T} \right|,$$

where

- V_α is rectangle window centered at a fixed point and turned on the angle α from the horizontal position,
- natural T and K represent, respectively, the length of a row and the number of rows of V_α as on Fig. 1B.
- $f_{V\alpha}(t)$ is the function of one variable defined on the integers of the interval $[1, KT]$, which unfolds image inside V_α. Formally, $f_{V\alpha}$ is defined by equation $f_{V\alpha}(t) = I_{V\alpha}(u,w)$ where $t = u + Tw$ and $\{u,w\}$, $u = 1,\ldots,T$, $w = 1,\ldots,K$, represents a node of the lattice in V_α.

Given a rectangle-wise window V_α with a fixed center $\boldsymbol{C}$ and a variable parameter of orientation α, the 3FM works as follows: *Find such value α^* of α for which $\{(\alpha^* = arg\ loc\ max\ S^0{}_{V\alpha})\ \&\ (S^1{}_{V\alpha^*} = S^l{}_{V\alpha^*} = 0)\}$ holds. If such α^* is found, 3FM signals detection of a gedgel centered at $\boldsymbol{C}$ with orientation α^*.*

See [9] for mathematical justification of the 3FM.

Post-processing filter of 'slope-wise' gedgels. (It is an extension of [9]). Detection of some objects covered by Definition 1 can be undesirable in some situations.

In particular, let us assume that image I is constructed just as a 'slope-wise' brightness, that is, as an arbitrary linear function of two arguments with non-zero gradient. For such I, due to Definition 1, a gedgel orthogonal to the gradient of I will be detected at any point of the image frame. Detection of such objects can be desirable or not, depending on application context.

A post-processing filter was introduced after 3FM to control acceptance/rejection of the 'slope-wise' gedgels. The idea of this filter is as follows. Let I_V be a gedgel represented by a positive response of 3FM. Using the same local coordinates as in Definition 1, I_V is subjected to a standard statistical test (verification of 0-hypothesis [7]) on the hypothesis that the correct functional model for mathematical expectation of I_V is given by the function $\lambda+\mu u$ with some real λ and μ. Some thresholds controlling the test vary the share of 'slope-wise' gedgels to pass the filter. Further details of the 'slope-wise' filter are omitted here.

Fig. 2 illustrates action of the filter.

3. Multiresolution Image Sketching

The term `multiresolution sketch` (MS) will stand for a hierarchical structure constructed by our image processing. Decomposition of the MS into smaller parts is presented in definitions of Table 1.

Some details are commented below.

While constructing a slice image I_s, the transform L_s works practically as image defocusing with Gaussian smoothing.

The 3FM combined with 'slope-wise' filter was used as the detector for construction of TCs.

In our experiments, while constructing the MS, the following parameters were varied in a coordinated way: L_{Si}, the scale S_i, and parameters of 'slope-wise' filter. Practically, this coordination followed the rule that the stronger defocusing of a slice image, the greater S_i, and the greater share of gedgels allowed to pass the slope-wise filter.

[In experiments, the sizes of lattice of *TW* were fixed as K=16, T=12. The ith slice image is constructed as convolution of the original with a Gaussian bell-wise matrix of weights. The size N_i of the ith matrix of weights is related to scale S_i by equation $S_i = (K+ N_i - 1)/K$, where $N_0=\sqrt{\{size\ of\ an\ image\ side\}}$, and for $i\geq 0$, recursively, $N_{i+1}=\max\{N_i/1.2;1\}$. The construction of slices for MS stops at $N_i=1$. The 'slope-wise' filter allows passing a greater share of slope-wise gedgels for low values of i. With growth of i, the filter rejects more and more slope-wise gedgels.]

Table 1. Architecture of MS and auxiliary definitions

Test window (*TW*) is rectangle $V_{x,y\alpha}$ with geometrical sizes $SK\times ST$ and a regular $K\times T$-lattice ($K>T$) inside; The $\{x,y\}$ and α represent, respectively, the center of *TW* and the angle between its long side and the horizontal axis;

The size S of square sells of *TW* is called **scale**.

Gedgel configuration space is 4D manifold $F\times J\times R_+$, where $F\subset R^2$ is image frame, J is interval $[0, \pi]$ with identification $0=\pi$ of its ends; Components of F, J, and R_+, represent, respectively, location, direction, and scale of gedgel. A point of this space is called **gedgel (**see *Terminology note* of Section 2). A gedgel is represented as quadruple $\{x,y,\alpha, S\}$, where $\{x,y\}\in F$, $\alpha\in J$, and $S\in R_+$; the first three components $\{x,y,\alpha\}$ describe a TW $V_{x,y,\alpha}$, S represents its scale.

Characteristic domain of gedgel $\{x,y,\alpha, S\}$ is the test window $V_{x,y,\alpha}$ with scale S, reconstructed by the gedgel parameters.

Tracked curve (TC) is an ordered set of gedgels with a common scale, obtained by a bi-directional tracking which starts from a nucleus gedgel.

Slice object (SO) is a set of TCs constructed for the same scale by the following sewing rule: two TCs, C_1 and C_2, belong to a common SO iff there exist such two gedgels $g_1\in C_1$ and $g_2\in C_2$, that are sufficiently close for a certain metric in the gedgel configuration space.

Slice image $I_S = L_S(I)$ is an image constructed from the original image I, subjecting it to a transform L_S from a one-parameter family $\{L_p\}_{p\in R}$ of transforms.

Slice Σ_S is a data structure formed as union of all SOs constructed by a slice image I_S.

Multiresolution sketch (MS) is a data structure formed as union of totally ordered set of slices $\Sigma_{S0} \prec \Sigma_{S1} \prec ... \prec \Sigma_{SN}$ with the order induced by values of scales $S_0 < S_1 < ... < S_N$.

4. Singular Gedgels as New Features and Matching Them in Different Shots

Table 2 contains definition of robust singular gedgels (SGs) and inter-slice objects (ISOs). The algorithm of their construction is presented in Table 3.

SG is designed for detecting in the MS a local geometric invariant of the following kind (see Fig. 1C): Let X be a half transparent smooth film in R^3 that contains a vertical line V which crosses the ground plane at the point $\{x,y\}$, and let the family $\{C_h\}$ of horizontal curves is formed as intersections of X with a family of horizontal planes $\{P_h\}$ scaled by the level h from the ground plane. Imagine the synthetic image as the shade that X throws onto horizontal plane when the sun is at zenith. It is clear that if $\{C_h\}$ has a considerable torsion around V then there is a stable ellipsoid of the most obscure shade centered at $\{x,y\}$.

Table 2: Definitions used in Section 4

Synthetic image of gedgel is defined in the gedgel characteristic domain as 2D normal distribution normalized with respect to gedgel scale. More exactly, it is equal to $\frac{1}{2\sigma_1\sigma_2} e^{-\frac{u^2}{s^2\sigma_1^2}-\frac{w^2}{s^2\sigma_2^2}}$, where u,w are such coordinates in the own gedgel coordinate system (like in Fig.1A) that $(u,w)=(0,0)$ is the centre gedgel, S is the gedgel scale, σ_1 and σ_2 are some parameters.

Synthetic image of MS is sum of synthetic images of all gedgels of this MS.

Characteristic function τ is a function defined on the set of gedgels of MS; In this paper, $\tau(g)$ is absolute value of the first or, optionally, the second derivative of synthetic image of MS computed at the centre of g along the only TC that contains g.

Singular gedgel (SG) g^* is defined as $\arg\max_{g\in G} \tau(g)$, where G is a variable subset of gedgels of MS.

Inter-slice tolerance of gedgel *g* is a neighborhood D_g in gedgel configuration space. It is designed to decide, whether or not a gedgel of arbitrary slice of MS is sufficiently close to a gedgel g of a certain slice.

Inter-slice object (ISO) **generated by SG** g is union of those SOs of different slices of MS that contain at least one gedgel of inter-slice tolerance D_g.

Given pair $\{r, s\}$ of SGs, **transform *T* induced by {*r, s*}** is an endomorphism of gedgel configuration space (see the main text for more details on T).

Similarity measure for two SGs {*r , s*} of different shots is pair $\{ N_r^s/|G_r|, |G_r| \}$ where $|G_r|$ is the total number of gedgels in ISO generated by r and N_r^s is the number of such gedgels g of G_r that $T(g)$ is sufficiently close to a gedgel of G_s, where T is the transform induced by $\{r , s\}$.

Construction of SGs and ISOs is based here on the use of a synthetic image constructed with the MS. Traditionally, the synthetic image serves only for qualitative estimation of an image processing output. So now, it obtains a new role.

Our scheme of construction of synthetic image is based on a straightforward intuitive idea determined formally in Table 2. See [14] for an advanced approach to image reconstruction by detected features. Examples of synthetic images are presented in Fig. 2.

Table 3: Algorithm for calculating ISOs

Inputs: $1>>\delta>0$, MS, synthetic image, characteristic function $\tau(g)$.
Returns set Ψ of all ISOs.
begin Initialize: Ψ to $\varnothing$, set G to all gedgels of MS,
m^* to $\max_{g\in G} \tau(g)$;
while $(G\neq\varnothing \wedge \max_{g\in G} \tau(g)>\delta m^*)$, **repeat**:
{ Construct $g^*\in G$ as a new SG;
Delete from G g* and the gedgels adjacent to g* in the only TC that contains g*;
Initialize new ISO ϑ to the SO that contains g^*;
Construct set $\{g_i\}$ with all gedgels of MS that belong to inter-slice tolerance D_{g^*};
For any $g\in \{g_i\}$, extend ϑ by gedgels of SO that contains g; delete from G g and the gedgels adjacent to g in its own TC;
Add ϑ to Ψ. }
end

The transform T induced by a pair of gedgels $\{r, s\}$ (see Table 2) is designed as a tool for solving the correspondence problem. Under assumption that r and s represent the same detail of a real object, but for two different shots, then, given gedgel g of ISO G_r generated by r, T transfers g to gedgel $T(g)$ (as a point in configuration space) with a high expectation to contain in its close neighborhood a (really detected) gedgel of ISO G_s generated by s.

This property of T is exploited in Table 2 in the definition of similarity measure. The first component in the pair $\{ N_r^s/|G_r|, |G_r| \}$ is properly the similarity measure; the second one is needed to control the confidence: the larger value of $|G_r|$ the more confident (statistically) $N_r^s/|G_r|$.

A simplified analytical form for induced transform T is as follows: Let $r=\{x_r, y_r, \alpha_r, S_r\}$, $r\in G_r$, $s=\{x_s, y_s, \alpha_s, S_s\}$, $s\in G_s$, and let $g=\{x, y, \alpha, S\}$ be arbitrary gedgel of G_r. Let $\Delta\alpha$ be defined as $\Delta\alpha=\alpha_s-\alpha_r$, $R_{\Delta\alpha}$ be matrix of rotation on angle $\Delta\alpha$, vectors $\mathbf{w}, \mathbf{u}\in R^2$ be defined as $\mathbf{w}=\{x, y\}-\{x_r,y_r\}$ and $\mathbf{u}=\frac{S_s}{S_r} R_{\Delta\alpha}\mathbf{w}+\{x_s,y_s\}$, respectively. Let $\{x',y'\}$ be coordinate decomposition of $\mathbf{u}$, α' and S'

be defined, respectively, as $\alpha+\Delta\alpha$ and $\frac{S_s}{S_r}S$. In this notation, one has $T(g)=\{x',y',\alpha',S'\}$. I.e., T acts as a combination of rotation, re-scaling and translation, the same for all gedgels of G_r .

This representation of T looses adequateness with growth of diameter of G_r, if a real object, which r corresponds to, is subjected to rotation around an axis parallel to the image plane. Nonetheless, it has turned out sufficiently good for the experiments below.

5. Consistency

Table 4 presents a RANSAC-based algorithm that establishes correspondences between singular gedgels extracted from two shots of a film.

Table 4: An algorithm for calculating consistent groups of correspondences

Inputs: N, h, m_{size}; parameters (commented in the main text) used in steps 2 and 3: parameters of ε- and ε_1-cells, radii of rings R and R_b, thresholds d_1, d_2 .
Returns the union M_* of consistent groups of pairs of corresponding elements.
Initialize M_* to $\varnothing$, U_1 and U_2 to full sets of SGs in the first and the second shot, respectively.
repeat steps 1-6 N times:

1. Determine $U \subset U_1 \times U_2$ as all pairs of gedgels $\{r, s\}$, $r \in U_1$, $s \in U_2$, with distance between their $\{x,y\}$-components less than a threshold h;
2. Determine a seed correspondence $u \in U$.
3. Extend u to a maximum set $M_u \subset U$ under constraints on local variations of correspondences.
4. Go to step 2, if the number of elements in M_u is less than threshold m_{size}
5. Reduce U_1 and U_2 by the projections of M_u on the first and second components of pairs, respectively.
6. Add M_u to M_* .

The term '*ε-cell*' stands for a neighborhood of a (small) fixed size (in a certain metric) of the transform induced by a correspondence of two gedgels (Table 2).

The step 2 works as follows. Choose at random center of a small ring R in the first shot. Construct ε_R as an ε-network for the set $E_R \subset U$ of potential pairs of corresponding elements determined by the condition that the first components of pairs belong to R. (Hence, the induced transform of any elements of E_R belongs to an ε-cell of the induced transform of a node of ε_R.) Construct a histogram with nodes of ε_R as the arguments whereas the value at a node to be a weighted sum of the elements of E_R whose induced transforms belong to ε-cell of the node. Use the values of similarity measure as the weights. At last, choose $u \in \varepsilon_R$ as *arg max* of the histogram.

The step 3 works as follows. Initialize M_u to the pair represented by u and B_M (the 'current border' of M_u) to the set of all pairs that fall into the same peak of the histogram as u. Then, up to stabilization of M_u, do:

{Add B_M to M_u. Construct a new histogram for each element $b \in B_M$ regarded its first component as the center of ring R_b. However, now apply this histogram in a different way: choose not global, but a local *arg max* from the nodes that belong to ε_1-cell of b. Actualize B_M to set of such local maxima for all b.}

The thresholds d_1, d_2 are used, respectively, to reject the seed u if B_M at the first iteration is too small, or the first iteration gives a too small increment to M_u.

Remark 2. Notice that (*i*) ε_1-cell represents a constraint on local variation of Jacobian of smooth homeomorphism that matches images. (*ii*) Construction of the histogram is similar to the Hough transform, but constructed for the configuration space of the pairs of (potentially corresponding) gedgels.

6. Experiments

Fig.2 gives an idea of what an MS can tell us about the original, and the influence of slope-wise filter. Fig.3 illustrates main components of MS.

Figure 2. Influence of slop-wise filter. From left to right: the original and three synthetic images that correspond to the gradually incremented allowed share of slope-wise gedgels. The numbers of detected SGs are, respectively, 700:1400:2800, approx.

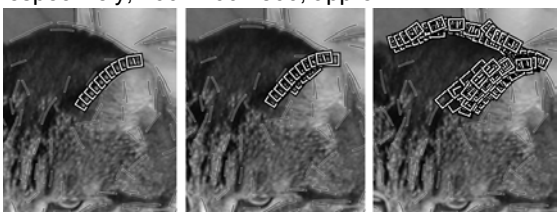

Figure 3. Components of MS. From left to right: samples of a TC, a slice object, and an ISO, respectively, constructed in the same area of the image of Fig.2

Fig. 4-5 present output of the algorithm of Table 4 for shots from a 'kitchen scene' clip with a considerable variation of image depth, active environment and camera viewpoint. The time lag between two shots of Fig.4 is 0.5s. There were dynamic occlusions and articulated objects of a complex shape. The maximum

shift of correspondent elements in the shots was about 17% of the frame diagonal.

Precision of found correspondences is respectively low (2-4 pixels) compared to photogrammetry methods (a fraction of pixel). A 'correct' M_u may contain a few mismatches, especially related to moving occlusions and spatial borders between different objects. A small percent of groups of correspondences M_u is wrong. Fig. 6 shows an example of such a wrong M_u found in an experiment.

Figure 4. Two shots of a 'kitchen scene' clip with all found correspondences in an experiment. For convenience of understanding the results, the same icon for any matching pair was chosen at random from a limited set of options. The number of SGs for each shot was about 900. The number of found matching pairs is 196.

7. Conclusion

Experiments have shown that the network of singular gedgels constructed by the developed technique contains a representative subnet, which can be efficiently used in applications. It is the matter of future work to determine the range of applicability.

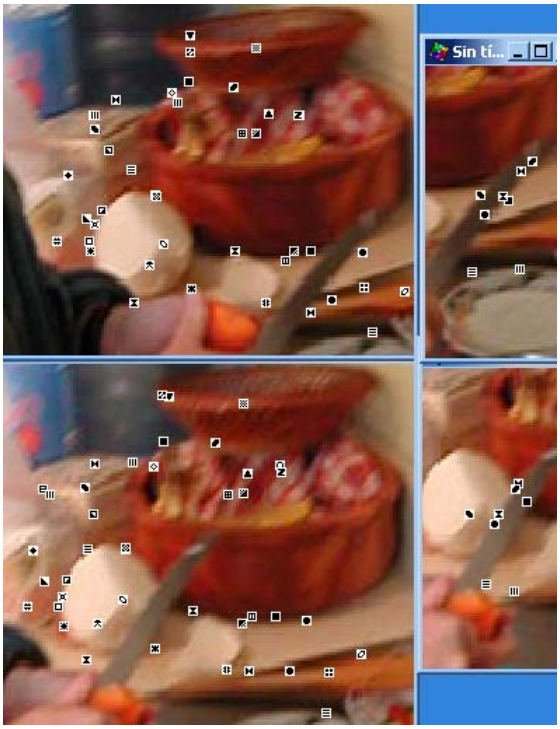

Figure 5. Either pair of image fragments contains just one group of pairs of corresponding elements (an M_u in terms of Table 4) found in experiments with the same shots as in Fig.4. The right M_u 'sees' only the knife. The left M_u 'sees' background, but 'does not see' the knife.

Figure 6. A wrongly detected group of correspondences

In this respect, consider Fig. 7 where unmodified algorithm of Section 5 was applied not to the shots of a film, but to a couple of face pictures. A tool for automatic matching face details of two arbitrary faces, ignoring semantic of the details, would be extremely useful for the face recognition and analysis. The results presented on Fig. 7 are promising, but not satisfactory yet for practice.

Notice, however, that in the algorithm of Table 4, we intentionally avoided using epipolar constraints to stress own potential of the new features; it was designed to deal with a dynamic scene and occlusions; it involves only a part of available information because it explores not geometrical, but only metrical constraints; it does not reject mismatches that lead to the local folds which are not possible in reality. These arguments allow us to assert that a specialization of the algorithm can improve significantly this preliminary example of automatic matching of face details.

Figure 7. Automatic matching face details: A straightforward application of the same algorithm as in Fig.4-5, but for two different portraits.

The presented approach should be studied and optimized in many respects: the numerical complexity, alternative early-vision detectors, the number of slices, using alternative kinds of geometric singularities for construction of singular gedgels, to mention just a few.

Although precision of SGs is not high, measurement of the group dynamics of SGs can be useful for scene understanding, learning object models, and robotics.

These are some challenging research lines in further development of the approach.

Acknowledgement

This work was supported by CONACYT under grant 400200-5-34812-A.

References

[1] J.F. Canny. A computational approach to edge detection. *PAMI*, 8(6):679-698, November 1986.

[2] M. Fischler and R. Bolls, Random sample consensus: A paradigm for fitting with application to image analysis and automated cartography, Communication of the ACM **24**(6), 381-395, 1981.

[3] W.Förstner, Image Matching, Chapter 16 in R. Haralick and L. Shapiro, *Computer and Robot Vision*, Vol.2, pp. 289-378, Addison-Wesley,1993.

[4] D.A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach,* Reading, Prentice Hall, 2002.

[5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[6] M.Hueckel, An operator which locates edges in digital picture, *JACM*, 18, 1, pp.113-125, January 1971.

[7] R.A. Johnson, *Miller and Freund's Probablitiy and Statistics For Engineers*, 5th Ed., Prentice-Hall, 1997.

[8] M. Kass, A.Witkin, and D. Terzopoulos, Snakes: Active Contour Models, *Proc. Int'l Conf. on Computer Vision*, pp. 259-269, 1987.

[9] G. Khachaturov, An Approach to Detection of Line Elements, *Proc. of the Second Asian Conference on Computer Vision (ACCV'95),* Vol. III, pp.559-563, 5-8 December 1995, Singapore, 1995.

[10] T. Lindeberg. Edge Detection and Ridge Detection with Automatic Scale Selection*. IJCV*, 30(2), 117-156, 1998.

[11] D. Marr and E.Hildreth, Theory of Edge Detection, in *Proc. Royal Society of London,* B-207, pp. 187-217, 1980.

[12] D. Marr and T.Poggio, A computational theory of human stereo vision, *Proc. Royal Society of London,* B-204, pp. 301-328, 1979.

[13] J. Matas, O. Chum, U. Martin, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions, *Proceedings of the British Machine Vision Conference*, volume 1, pp. 384-393, London, 2002.

[14] V.S. Nalwa, *A guided tour to computer vision*, Addison-Wesley,1993

[15] Ch. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity,* Reading, Dover Publications, 1998.

[16] W.K. Pratt, *Digital image processing*, Reading, John Wiley &Sons, Inc., 1978

[17] M.A. Ruzon and C. Tomasi, Edge, Junction, and Corner Detection Using Color Distribution, *PAMI*, Vol.23, 11, pp. 1282-1295, 2001.

[18] S.M. Smith and J.M. Brady, SUSAN - a New Approach to Low Level Image Processing, *IJCV,* 23(1):45-78, May 1997.

[19] M. Sonka, V. Hlavac, R. Boyle, *Image processing, analysis, and Machine Vision*, Reading, International Thomson Publishing, 1999.