A Fast Probabilistic Model for Hypothesis Rejection in SIFT-Based Object Recognition*

Patricio Loncomilla and Javier Ruiz-del-Solar

Department of Electrical Engineering, Universidad de Chile

Abstract. This paper proposes an improvement over the traditional SIFT-based object recognition methodology proposed by Lowe [3]. This improvement corresponds to a fast probabilistic model for hypothesis rejection (affine solution verification stage), which allows a large reduction in the number of false positives. The new probabilistic model is evaluated in an object recognition task using a database of 100 pairs of images.

1 Introduction

Object recognition approaches based on local invariant features have become increasingly popular and have experienced an impressive development in the last years ([1][3][5][8][11]). Typically, local invariant features are extracted from a test image, then characterized by invariant descriptors and finally matched against a reference database. Most employed local detectors are the Harris detector [2] and the Lowe's sDoG+Hessian detector [3], being the Lowe's detector multiscale and the Harris detector single scale. Best performing affine invariant detectors are the Harris-Affine and the Hessian-Affine [10], but they are too slow to be applied in general-purpose object recognition applications. The most popular and best performing invariant descriptors [9] are the SIFT (Scale Invariant Feature Transform) features [3].

When building real-world object recognition applications as for example robot self-localization systems based on invariant visual landmarks [12] or robot head pose detection systems [6], the algorithm recognition capabilities and processing speed are both important. Lowe's system [3] using the sDoG+Hessian detector, SIFT descriptors and a probabilistic hypothesis rejection stage has acceptable recognition capabilities and works in near real-time (1-3 images per second). However, Lowe's system main drawback is the large number of false positive detections. This is a serious problem when using it in vision tasks that need to process video sequences of images.

For that reason, the aim of this paper is to improve the traditional SIFT-based object recognition method from Lowe, by proposing a fast probabilistic model for hypothesis rejection (affine solution verification stage), which allows a large reduction in the number of false positives. The new probabilistic model is evaluated

* This research was partially supported by FONDECYT (Chile) under Project Number 1061158.

in an object recognition task using a database of 100 pairs of images (UCH100 database).

This article is structured as follows. In section 2 we describe the proposed fast probabilistic model for hypothesis rejection. Experimental results of applying this probabilistic model in the recognition of objects present in real work images (UCH100 database) are presented in section 3. Finally, in section 4 some preliminary conclusions of this work are given.

2 Fast Probabilistic Model for Hypothesis Rejection

As already mentioned Lowe's system use the sDoG+Hessian detector, SIFT descriptors and a probabilistic hypothesis rejection stage. The system is very complex, having several sub-stages (local extrema detection, accurate keypoint localization, orientation assignment, etc.). A detailed description can be found in [3].

One of the main weaknesses of Lowe's algorithm is the use of just a simple probabilistic hypothesis rejection stage, which cannot successful reduce the number of false positives. Lowe's method for calculating a probabilistic model for hypothesis rejection [4] requires that the explicit affine transformation must be known in advance, and that all matches that fall onto the projected region must be counted. Given that the probabilistic model is applied after the matching stage, all bins with more than 4 votes must be full-processed. This computation can slow down the process if the number of bins and matches is large.

In this section an additional fast probability rejection test is proposed. It consists on assigning a probability value to all bins with more than 4 votes, without knowing an explicit transformation. This probability values are calculated directly in the quantized Hough bins space. This allows the rejection of bins with very low probability without the requirement of additional processing.

A general similarity transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ has the following expression:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} e\cos\theta & e\sin\theta \\ -e\sin\theta & e\cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_X \\ t_Y \end{bmatrix} \Leftrightarrow \mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{t} \tag{1}$$

$\mathbf{A},\mathbf{t}$ depend on the $(\Delta x, \Delta y, \Delta\theta, \Delta n)$ differences of the object pose between the two compared images, Δn being the differences in the scale dimension.

A similarity transformation that quantizes the pose difference $(\Delta x, \Delta y, \Delta\theta, \Delta n) = (x_1 - x_2, y_1 - y_2, \theta_1 - \theta_2, n_1 - n_2)$ in bins of size $\left(\frac{L_X}{4}, \frac{L_Y}{4}, 30^\circ, 1\right)$, as a function of integer variables (i, j, k, z), has the following expression [3]:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = 2^Z \begin{pmatrix} \cos(30^\circ k) & \sin(30^\circ k) \\ -\sin(30^\circ k) & \cos(30^\circ k) \end{pmatrix} \begin{pmatrix} x - 0.25 L_X i \\ y - 0.25 L_Y j \end{pmatrix} \tag{2}$$

From (1) and (2) we obtain fractional values for (i, j, k, z):

$$
\begin{aligned}
i_{FRAC} &= \frac{2^{\frac{n_1-n_2}{2}} x_2 - x_1\cos(\theta_1-\theta_2) + y_1\sin(\theta_1-\theta_2)}{(1/4)L_X \times 2^{\frac{n1-n2}{2}}} \\
j_{FRAC} &= \frac{2^{\frac{n1-n2}{2}} y_2 - y_1\cos(\theta_1-\theta_2) - x_1\sin(\theta_1-\theta_2)}{(1/4)L_Y \times 2^{\frac{n1-n2}{2}}} \\
k_{FRAC} &= \frac{\theta_1-\theta_2}{30^o} \\
z_{FRAC} &= \frac{n_1-n_2}{2}
\end{aligned}
\tag{3}
$$

Each match $(x_1,y_1,\theta_1,n_1)\leftrightarrow(x_2,y_2,\theta_2,n_2)$ has 16 nearest values (i,j,k,z) for which it must vote. It can be observed that (i,j) quantizes translation, k quantizes rotation and z quantizes scale difference in the similarity transformation.

The probability that a random incorrect match votes for a given bin $B=(i,j,k,z)$ in the bin-space is $p_B = p(i,j,k,z) = p(z)p(k)p(i,j\,|\,k,z)$. When a correct mapping m_B for the bin B does not exist, all the votes in bin B are of random origin. Each random match votes for the 16 different nearest bins. We can estimate the probability that k or more random incorrect matches of a total of n vote for a bin B (cumulative binomial distribution):

$$
P(\#B \geq k\,|\,\neg m_B) = \sum_{\alpha=k}^{N} \binom{N}{\alpha} {p_B}^{\alpha} (1-p_B)^{N-\alpha} \tag{4}
$$

with $\#B$ the number of votes in the bin B and $N=16\times n$ the total number of random votes generated by the n matches that exists in all the bin-space. This approximation is acceptable when k is much smaller than n, as each random match produces 16 random (distinct) votes.

The probability of a bin B representing a true mapping m_B of an object can be approximated as [4]:

$$
P(m_B\,|\,\#B \geq k) = \frac{P(m_B)}{P(m_B) + P(\#B \geq k\,|\,\neg m_B)} \tag{5}
$$

An exact value of $p_B = p(z)p(k)p(i,j\,|\,k,z)$ is essential for obtaining an exact computation of (4) and (5). Lowe approximates $p(z)=0.5$. But, if it is assumed that the density of interest points along the sub-sampled scale space is constant, an analytical value for $p(z)$ exists and can be computed. Lowe also approximates $p(i,j\,|\,k,z)$ as a fixed value. But, $p(i,j\,|\,k,z)$ can be estimated as a ratio between the space covered by the matches compatibles with the bin (*i*,*j*,*k*,*z*) and the space covered by all the possible matches that can be generated between the pair of images. Finally, the probability $p(k)$ can be calculated as *w*/360°, where *w* is the angular width of a bin.

2.1 Analytical Computation of $p(z)$

Suppose we have a pair of images I and I'. $\{D_0, D_1, ...\}$ and $\{D'_0, D'_1, ...\}$ will be their respective sub-sampled scale space representations, and two images per octave will be used. The area of a scale space image can be expressed as:

$$area(D_k) = \frac{area(D_0)}{4^{floor(k/2)}},\ area(D'_k) = \frac{area(D'_0)}{4^{floor(k/2)}}$$

If the density of interest points is constant in each of the scale spaces, and the point-matches are of random origin, the probability that a random match will be associated to a (m,n) scale space level can be written as:

$$P(match : m \to n) = \frac{area(D_m)area(D'_n)}{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} area(D_i)area(D'_j)}$$

If we simplify the last expression, a simple analytical probability expression can be obtained:

$$P(match : m \to n) = \frac{9}{64}\left(\frac{1}{4}\right)^{floor(m/2)}\left(\frac{1}{4}\right)^{floor(n/2)}$$

It will be defined a Z function that depends on m,n: $Z(m,n) = FLOOR(m/2) - FLOOR(n/2)$. Now, the following set can be evaluated:

$$B(z) = \{(m,n) \mid Z(m,n) = z\}$$
$$= \begin{cases} \{(2k, 2k+2z), (2k+1, 2k+2z), (2k, 2k+1+2z), (2k+1, 2k+1+2z) \forall k > 0\}, z > 0 \\ \{(2k-2z, 2k), (2k+1-2z, 2k), (2k-2z, 2k+1), (2k+1-2z, 2k+1) \forall k > 0\}, z < 0 \end{cases}$$

Using this set, probabilities in the (m,n) space can be mapped to the z space.

$$p(z \mid z \geq 0) = \sum_{k=0}^{\infty} P(match : 2k, 2k+2z) + \sum_{k=0}^{\infty} P(match : 2k+1, 2k+2z) + \\ + \sum_{k=0}^{\infty} P(match : 2k, 2k+2z+1) + \sum_{k=0}^{\infty} P(match : 2k+1, 2k+2z+1) \quad (6)$$

$$p(z \mid z < 0) = \sum_{k=0}^{\infty} P(match : 2k-2z, 2k) + \sum_{k=0}^{\infty} P(match : 2k+1-2z, 2k) + \\ + \sum_{k=0}^{\infty} P(match : 2k-2z, 2k+1) + \sum_{k=0}^{\infty} P(match : 2k+1-2z, 2k+1) \quad (7)$$

Finally, (6) and (7) can be simplified to:

$$p(z) = \frac{3}{5}\left(\frac{1}{4}\right)^{|z|}$$

It can be easily demonstrated than $\sum_{z=-\infty}^{\infty} p(z) = 1$. The new $p(z)$ expression can be used to get a modified probability test to reject incorrect bins.

2.2 Analytical Computation of $p(i, j \mid k, z)$

The $p(i, j \mid k, z)$ calculation considers the space of all positions (x_1, y_1) in the test image and all the (x_2, y_2) positions in the reference image. Random matches between the images generate random (x_1, y_1, x_2, y_2) points in a 4D space. If the reference image size is $L_X \times L_Y$ and the test image size is $M_X \times M_Y$, the 4D random points belong to the following space:

$$\mathbf{c} = (x_1, y_1, x_2, y_2) \in S = [0, M_X] \times [0, M_Y] \times [0, L_X] \times [0, L_Y]$$

The S space has a 4D volume that can be expressed as $L_X L_Y M_X M_Y$. A bin B covers a subset of S that will be named $Q(B)$. If the 4D volume contained by $Q(B)$ is known, the probability $p(i, j \mid k, z)$ can be estimated as:

$$p(i, j \mid k, z) = \frac{Q(i, j, k, z)}{L_X L_Y M_X M_Y}$$

The last equation can be approximated and written in terms of the (i, j) space instead of the (x_1, y_1, x_2, y_2) space. We will analyze 3 cases:

Case 1: If we assume that $\Delta\theta = 0°$ in (3), the equations for i_{FRAC} and j_{FRAC} are reduced to:

$$i_{FRAC} = \frac{2^z x_2 - x_1}{(1/4) L_X \times 2^z}$$

$$j_{FRAC} = \frac{2^z y_2 - y_1}{(1/4) L_Y \times 2^z}$$

The minimum and maximum admissible values for i_{FRAC} and j_{FRAC} while (x_1, y_1, x_2, y_2) belongs to S are the following.

$$i_{FRAC} \in \left[\frac{M_X}{(1/4) 2^z L_X}, \frac{2^z L_X}{(1/4) 2^z L_X} \right], j_{FRAC} \in \left[-\frac{M_Y}{(1/4) 2^z L_Y}, \frac{2^z L_Y}{(1/4) 2^z L_Y} \right] \quad (8)$$

We define the following variables.

$$R_{XX} = \frac{M_X}{2^z L_X}, R_{YY} = \frac{M_Y}{2^z L_Y}$$

Then, expression (8), which expresses the domain for (i, j), can be rewritten as:

$$(i_{FRAC}, j_{FRAC}) \in [-4 R_{XX}, 4] \times [-4 R_{YY}, 4]$$

All the (i, j) bins have size 1 in the (i_{FRAC}, j_{FRAC}) space. Them the probability that a random (x_1, y_1, x_2, y_2) match produces a random (i_{FRAC}, j_{FRAC}) which vote for a particular (i, j) bin can be expressed as:

$$p(i,j \mid k,z) \approx \frac{1}{4(R_{XX}+1)\cdot 4(R_{YY}+1)}$$

Case 2: If we assume that $\Delta\theta = 90°$ in (3), a calculation similar to case 1 gives the following results:

$$R_{YX} = \frac{M_Y}{2^z L_X}, R_{XY} = \frac{M_X}{2^z L_Y}$$
$$p(i,j \mid k,z) \approx \frac{1}{4(R_{XY}+1)\cdot 4(R_{YX}+1)}$$

Case 3: If we do not assume a particular $\Delta\theta$ it is difficult to get an analytical solution. But an approximation can be obtained by mixing the two results. As R_{XX} and R_{YX} stands for two different orthogonal cases, they can be mixed as $R_X^2 = R_{XX}^2 + R_{YX}^2$. In a similar way, $R_Y^2 = R_{XY}^2 + R_{YY}^2$ can be assumed. This leads to the following equations:

$$R_X(\theta_1-\theta_2) = \frac{\sqrt{M_X^2\cos^2(\theta_1-\theta_2) + M_Y^2\sin^2(\theta_1-\theta_2)}}{2^z L_X}$$
$$R_Y(\theta_1-\theta_2) = \frac{\sqrt{M_Y^2\cos^2(\theta_1-\theta_2) + M_X^2\sin^2(\theta_1-\theta_2)}}{2^z L_Y}$$
$$p(i,j \mid k,z) \approx \frac{1}{4(R_X+1)\cdot 4(R_Y+1)}$$

3 Experimental Results

In this section is presented an experimental evaluation of the proposed improvement over Lowe's work. The performance of the introduced verification and merging hypothesis stages are tested in the UCH100 object recognition database (available in [13]). This database is composed by 100 pairs of real-world images $\{(I_{2k-1}, I_{2k}), k = 1,...,100\}$, being I_{2k-1} a reference image and I_{2k} the corresponding test image. Each reference image shows a different, single object. The same object appears in the corresponding test image, viewed under different conditions (position, view angle, partial occlusion, in-plane and out-of the-plane rotation). In the test images can also appear objects not included in the reference images. In figure 2 are shown some examples of reference-test images pairs.

Object recognition experiments were performed in all image's pairs $\{(I_j, I_k), k, j = 1,...,100\}$. The experiments consist on finding the mapping that relates each pair of images. A pair of images has a common object only in 100 of the 10,000 cases to be analyzed. In these pairs of images (100) the recognition methods generate a variable number of transformations (0 to 10, or even more in some cases), although ideally just one transformation should be obtained. For the proposed experiment, a

pair of images is *solved* when the transformation with the best priority, i.e. the highest probability value, is a good-mapping transformation, and almost all the associated point-matches are correct. The other obtained transformations are not analyzed. The algorithms are compared in terms of:

- DR (Detection Rate): DR is computed as the rate of correct best-priority transformations. Just one per image can be correct in the 100 pairs having a common object.
- FPR (False Positive Rate): FPR is computed as the rate of incorrect best-priority transformations. Just one incorrect transformation per image is added when incorrect objects are matched.
- DR/FPR Ratio: Ratio between correct and incorrect best-priority transformations.
- Mean PT: Mean Processing time for the matching and verification stages.

The algorithms under comparison are:

- *Lowe*: Lowe's recognition system without any improvement.
- *FastProb*: Lowe's recognition system using the fast probabilistic model for hypothesis rejection.
- *Lowe+OVS*: *Lowe* plus other verification and merging stages proposed in [7].
- *FastProb+OVS*: *FastProb* plus other verification and merging stages (see [7]).

Table 1. Comparative evaluation of the different algorithms. DR=Detection Rate. FPR=False Positive Rate. Mean PT: Mean Processing time for the matching and verification stages.

Algorithms	DR (%)	FPR (%)	DR/FPR Ratio	Mean PT [ms]
Lowe	41%	85.5%	0.48	21.56
FastProb	39%	78.3%	0.50	14.38
Lowe+OVS	44%	4.87	9.03	26.56
FastProb+OVS	49%	3.74	13.10	19.38

The comparative evaluation of these algorithms is displayed in table 1. As it can be observed, the new FastProb rejection test reduces the FPR from 85.5% to 78.3%, while keeping the DR in about 40%. More important, the time required for performing the matching and verification processes is reduced from 21.56ms to 14.38ms (about 33% reduction). However, the FPR is still too high and other verification stages are required (see detailed explanation in [7]). When using these additional stages together with FastProb (FastProb+OVS) the DR is increased to 49%, while the FPR is strongly decreased to just 3.74, achieving a DR/FPR ratio of 13.10. When using the Lowe's algorithm together with the additional verification stages (Lowe+OVS), DR increases to 44%, FPR decreases to 4.87, and the resulting

Fig. 1. Some examples of object recognition results that can be obtained with the new probabilistic model for hypothesis rejection

DR/FPR ratio is 9.03. Thus, FastProb+OVS achieves a DR/FPR ratio 45% higher than Lowe+OVS. That means that the proposed fast probability model for hypothesis rejection is essential for obtaining high recognition rates. Figure 1 shows some examples of the excellent object recognition results that can be obtained when using this new model.

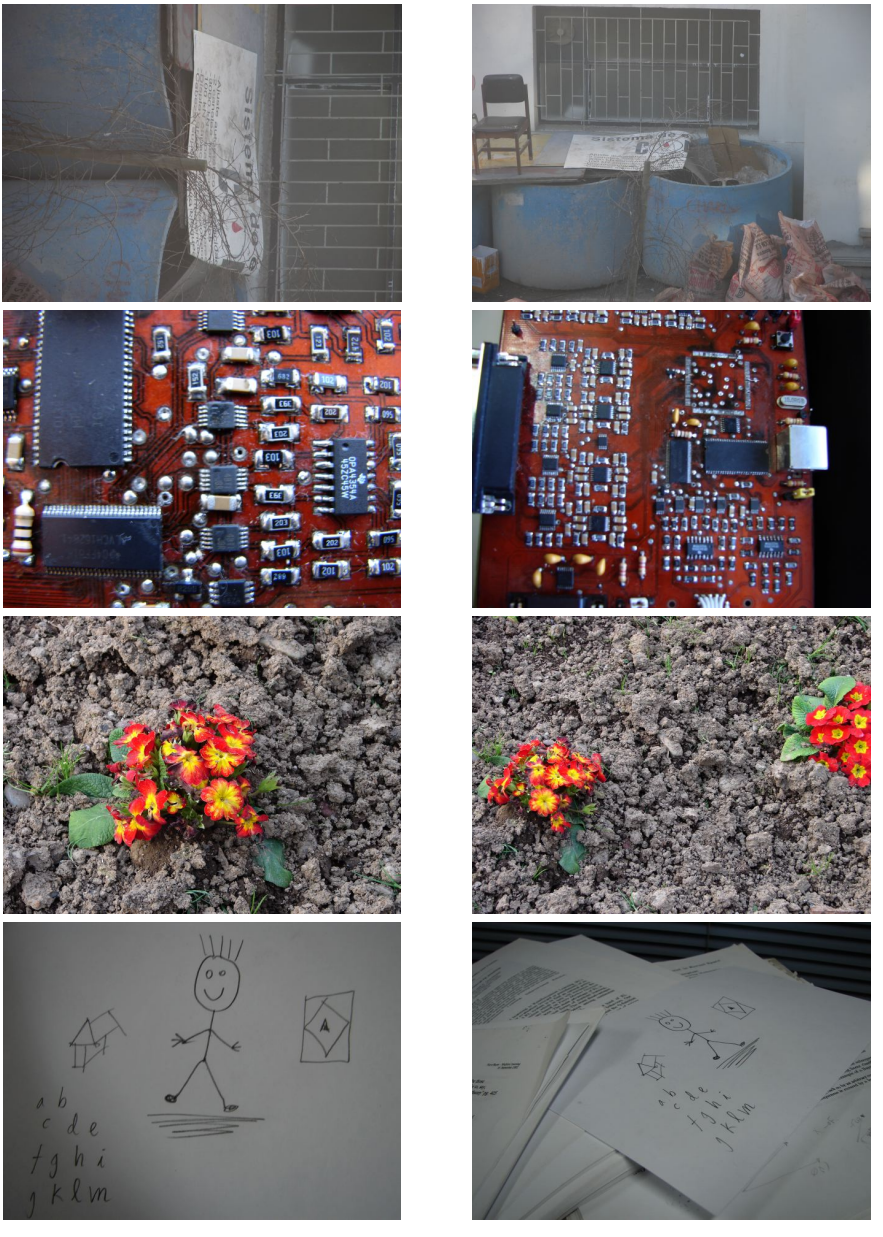

Fig. 2. Selected images from the UCH100 database (see [13]). Left: reference images. Right: corresponding test images.

4 Conclusions

In this work was proposed an improvement over the traditional SIFT-based object recognition methodology proposed by Lowe. This improvement corresponds to a fast probabilistic model for hypothesis rejection (affine solution verification stage), which allows a large reduction in the number of false positives. The new probabilistic model was evaluated in an object recognition task using a real-world database of 100 pairs of images. Objects in these images are very hard to recognize. The obtained results show that with the probabilistic model for hypothesis rejection is obtained a reduction in the number of false positives of about 9%, and the time required for the matching and verification processes is reduced in about 33%. This reduction is very important for several real-world applications.

References

1. V. Ferrari, T. Tuytelaars, and L. Van Gool. Simultaneous Object Recognition and Segmentation by Image Exploration. *Lecture Notes in Computer Science* 3021, 40 - 54.
2. C. Harris and M. Stephens, A combined corner and edge detector, *Proc. 4th Alvey Vision Conf.*, 147-151, Manchester, UK, 1988.
3. D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. Journal of Computer Vision*, 60 (2): 91-110, Nov. 2004.
4. D.G. Lowe. Local Features View Clustering for 3D Object Recognition. *Proc. of the IEEE Conf. on Comp. Vision and Patt. Recog*., 682 – 688, Hawai, Dic. 2001
5. P. Loncomilla and J. Ruiz del Solar. Improving SIFT-based Object Recognition for Robot Applications. *Lecture Notes in Computer science* 3617, Springer, 1084 - 1092.
6. P. Loncomilla and J. Ruiz-del-Solar. Gaze Direction Determination of Opponents and Teammates in Robot Soccer. *Lecture Notes in Computer Science* 4020, Springer, 230 – 242.
7. P. Loncomilla and J. Ruiz-del-Solar. An improved SIFT-based Object Recognition Methodology, *Tech. Report UCH-DIE-VISION-2006-03*, Dept. of E. Eng., U. de Chile, 2006.
8. K. Mikolajczyk and C. Schmid. Scale & Affine Invariant Interest Point Detectors. *Int. Journal of Computer Vision*, 60 (1): 63 - 96, Oct. 2004.
9. K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, 1615 – 1630.
10. K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool. A Comparison of Affine Region Detectors. *Int. Journal of Computer Vision* (accepted).
11. F. Schaffalitzky and A. Zisserman. Automated location matching in movies. *Computer Vision and Image Understanding* Vol. 92, Issue 2-3, 236 – 264, Nov./Dec. 2003.
12. S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *Int. J. of Robotics Research*, Vol. 21, No. 8, 2002, 735 – 758.
13. UCH100 database. Electronically available in: http://vision.die.uchile.cl/.